Does TAP have your Back?

An Impact Evaluation of an Organized Peer Tutorial Program in De La Salle University

Justin Raymond S. Eloriaga[1]
School of Economics, De La Salle University

Abstract

Organized Peer Tutorial Programs have proliferated in the past few years due to their ability to guide and encourage students to be able to perform better in academic endeavors. However, school management find it vital to obtain necessary external and internal validity measures to assess whether or not the tutorial programs are effective. This study seeks to determine the impact of the Economics Organization's Tutorial Aptitude Program (TAP) under the Office of the Associate Dean of the School of Economics on key outcome indicators in the academic performance of students studying in the university. A quasi-experimental methodology has been conducted using both a Two Stage Least Squares Framework and a Propensity Score Matching Framework to estimate various treatment effects on the effectivity of the program. It has been concluded that the TAP has had significant positive effects on the final grades of students suggesting that the program has met its goals of raising key academic indicators. However, a small sample and ethical considerations may have played a role in the insignificance of some estimates.

---

[1] Justin Raymond S. Eloriaga is the Research Assistant and Teaching Assistant under the Office of the Dean for the School of Economics as well as the Angelo King institute of Business and Economic Studies of De La Salle University. In addition, he is the 17th Young Economists' Convention Commissioner for the Economics Organization. All contact shall be done through the email justin_eloriaga@dlsu.edu.ph or at justin.eloriaga@icloud.com.

Table of Contents

Introduction

University life in the Philippine context has been influenced heavily by Western methodologies and methods. This particular life is very evident in students from Philippine universities who have very dynamic needs and schedules for college. The rigor of academics has always been the threshold of student life for almost any university (Paloyo et. al, 2016). Oftentimes, academic performance serves as a direct indicator of student productivity and is a benchmark which is widely considered by most corporations, firms and other hiring agencies[2]. Given the importance of academic performance, various educational institutions around the world (including the Philippine higher learning institutions) offer peer tutorial services whether subsidized by the university administration or organized by university student clubs and organization. These programs have increased in their popularity and serve as a new medium to reinforce academic knowledge gained with peer learning fostering great inter-student cooperation.

This study seeks to do an evaluation of such programs in the context of Philippine higher learning institutions (such as Philippine universities) and see whether these programs are effective in accomplishing their objective using an impact evaluation study. In particular, it shall evaluate the Tutorial Aptitude Program of the Economics Organization for a given time span in one subject.

Description of Intervention/ Project

Peer tutorial programs in the context of Philippine universities come in many varieties. Similarly, these programs have been adopted from various Western methods of peer learning with the goal of increased academic performance through peer learning. These programs operate under an assumption that students learning from their peers would better reinforce existing academic performance and will foster a medium of student exchange (Paloyo et. al, 2016)[3].

Peer tutorial programs may come in a variety of different forms, concepts or approaches with very similar underlying goals. Focus Group Discussions are common forms of peer tutoring centered around a question and answer format between tutees. Although quite informal, it is able to supposedly increase the academic performance of students, in particular, those who began with lower grades vis-a-vis higher grades[4]. Classroom approach is the most common of the approaches (in International settings) which makes use of a typical teacher to student format. Request approach is done through organizations with a pool of peer tutors who may serve as mentors. This setting is a more informal approach to the typical classroom

---

[2] Chances are, the higher the academic output or threshold, the more productive (behaviorally) a student is vis-a-vis other students.

[3] Peer learning is effective in the learning process and may lead to higher grades due to less of a generational gap and a more attuned experience overall.

[4] Similar to the evaluation of Paloyo et. al (2016), these types of discussions are similar to the peer evaluation programs present in their study which follow a format similar to an FGD or a colloquial termed "Alcohol Assembly" meeting

or traditional approach which makes use of fewer students as a whole. The three most common which appear in literature are the listed in table 1.

Table 1 Peer Tutorial Program Approaches

Approach	Description
Focus Group Discussion Approach	An informal discussion of various topics or questions with regards to a subject matter facilitated by a peer tutor. Usually, concepts discussed are just elaborations with no formal lessons and centers around question and answer for clarification purposes.
Classroom/ Traditional Approach	A more formal discussion which centers around a typical lecture session with possible accompanying exercises facilitated by a peer tutor. Concepts discussed would be very similar to the lecture room sessions but at a more controlled or modulated pace.
Request Approach	An approach that is by choice usually petitioned by students themselves or through recommendation from a department or administration. It follows a similar flow with the classroom or traditional approach but with less attendees per session due to a by request nature.

The approach implemented by different universities in the Philippines may vary per organization, department or administration. However, the underlying objectives of these programs may be aligned with a goal of increasing academic performance and standing of students in the university.

According to Burns (2006), peer tutorial programs have had significant effects on the academic performance of children, particularly those of special needs children who may be supplemented or encouraged to study better given a more relatable context[5]. This is further supplemented by a research conducted by Harper and Maheady in 2007 which said that students were often more engaged in the learning materials and were more adept to listen given teaching instruction from a peer vis-a-vis a regular class room style setting.

Statement of the Problem

Learning programs especially coming from peers are on the rise with regards to their popularity (Paloyo et al. 2016). While these programs are deemed helpful by students, it is necessary to determine their effectivity in an empirical and tried basis. Hence, this study focuses and revolves around determining the effectivity of such programs in the context. The main research problem that this study seeks to answer is whether or not peer tutorial programs in the context of a Philippine university is able to increase the academic performance of a student given various academic indicators.

---

[5] This relatable context being a peer vis-a-vis regular class room instruction

Objectives of the Evaluation

Peer tutorial programs may be programs which entail great effort from sponsoring administrations, organizations and departments. As such, it is important to be able to assess the effectivity of peer tutorial programs in the Philippine setting. Effectivity of these programs in the long run should be tested in order to improve or transform these initiatives to newer forms or refine them further if need be. The following is the central the objective of the study:

1. Determine the impact of peer tutorial programs on the academic performance of university students in the Philippines[6].

These two main objectives seek to identify the effectivity of these programs in an attempt to be able to recommend economically sound guidelines for administrators, organization and course departments present in Philippine universities. Note that while Philippine universities are the main treatment groups that would be present in the study, the actual study will only center around De La Salle University meaning localized effects would be found vis-a-vis general effects.

Significance of the Study

Peer tutorial programs are often sponsored by administrations, departments and organization which may entail great cost (Paloyo et. al, 2016). Programs such as these are often untested ideas and were implemented simply due to demand. Hence, they are new and often untested territory with results that may be biased to a given course or class. This study will seek to determine the effectivity of a particular peer tutorial program. The results of this study may be used by multiple administrations, and school organizations in order to improve the quality of their services for the betterment of the academic progression of various students in the university who may choose to attend or request peer tutoring services.

Scope and Limitations

Students part of the study will only come from Philippine universities taking up available courses offered in their respective colleges. Moreover, the study will only be conducted during a particular period in a given academic term. It will not span multiple time periods and will only be conducted for certain subjects wherein it is feasible to perform a more thorough evaluation due to changing standards and professors. These certain subjects include Mathematical Economics I, Mathematical Economics II, Econometrics 2 and Financial Econometrics typically taught in freshmen, sophomore or junior year for an economics major in the School of Economics in De La Salle University.

---

[6] Experiment is to be conducted on De La Salle University Junior and Senior Students which are students from a Philippine University

Impact Evaluation Fundamentals

This section shall give an overview of the general descriptions and expected outcomes of the impact evaluation study. It shall contain the hypothesis, the theory of change and the expected results change of the program in question.

Hypothesis

There may be a positive or stagnant effect in terms of peer tutorial program as the effectivity of these programs may vary across a number of different factors. Paloyo et. al (2016) notes that there was a noticeable improvement only on the students in the lowest bracket of the initial grade standing at a 90 percent confidence bound. However, other students of higher brackets did not show significant strides with regards to improvement at least statistically[7]. Moreover, other factors such as the adequacy of the peer tutors and the venue being conducive for learning must also be considered when making conclusions. With this, the hypotheses for this study are listed below.

H_1: Peer Tutorial Programs are effective programs in ensuring increases in academic performance of majority of the students who participated in the program.

H_0: Peer Tutorial Programs are not effective programs in ensuring increases in academic performance of majority of the students who participated in the program

The hypotheses listed seek to determine the breadth of effectivity or non-effectivity of peer tutorial programs given the indicators that will be stated in the methods. These indicators will determine or be correlated to factors which affect or define the academic performance of a student and other factors which may be correlated with other goals that are stipulated in the program.

Theory of Change

Peer tutorial programs are implemented in many ways highlighted by the different approaches that departments, colleges, administrations or organizations. There is an inherent process to how peer tutorial programs are organized and structured (McDuffie et al. 2009)[8]. This usually starts with a call for student tutors who would need to be trained for the job of a peer tutor. Next, the sponsoring body would determine the proper courses that would need to be accommodated based on the number of students who would be wanting a peer tutorial

---

[7] Statistically, results for higher initial brackets were not significant at a 90% confidence bound but were still treated as positive correlations despite very minuscule positive coefficients.

[8] These methods vary per the nature of a peer tutorial program given a predetermined procedure often deliberated by a sponsoring body

session. Lastly, the peer tutorial session would be conducted with a corresponding feedback of the session accomplished by the students. Table 2 summarizes this process below.

Table 2 Theory of Change

Process	Description	Expected Outcome or Deliverable	Causal Logic
Deliberation of Program Process	Concerns processes done by a sponsoring body with regards to planning a particular peer tutorial program (conventionally using one of the three main approaches discussed)	Project Plan and Action Plan as well as logistical concerns	Effect of the project design on the effectivity of the program
Call for Peer Tutors	Sponsoring body would interview and gauge the performance of potential peer tutors to see effectivity and adequacy in teaching a subject matter	Tutor list and tutor guidelines as well as training of interviewees to determine adequate tutors	Effect of the adequacy level of the tutor on the overall effectivity of the program
Training of Peer Tutors	Sponsoring body would train the peer tutors with regards to potentially assigned courses which may be offered		
Assigning of Courses for Tutors	Courses are assigned to tutors be it be through the demand of students or through the prerogative of departments and/or administration		

Peer Tutorial Proper	The program is conducted on a specified venue given a number of students for a particular course facilitated by a peer tutor	Lesson Plan and reviewers as well as frequently asked questions for cross feedback with faculty	Effect of the Peer Tutorial Program on Academic Performance
Feedback Delivery from Students	Feedback is to be received from students to better the program	Feedback Forms	
Feedback from Faculty on Student Academic Standing	Feedback by the faculty is given with regards to the current academic indicators of the students	Grade Standings and Examination Results	

Results Chain

Peer tutorial programs are programs which are implemented in many stages and will have certain outcomes and performance indicators as there is a progression. Hence, it is important to be able to establish the results chain in order to see a good overview of how a program will be implemented. This is the typical results chain of peer tutorial sessions and their corresponding inputs and outcomes. The results chain is listed in Table 3.

Table 3 Results Chain

Inputs	Activities	Outputs	Intermediate Outcomes	Final Outcomes
• Initial budget for the pilot randomized encouragement trial. • Initial incentives present for the randomized encouragement design.	• Training of peer tutors and available administrators on logistic processes for the programs • Venues to conduct the peer tutorial programs such	• Successfully conducted peer tutorial sessions • Transcription of attendance on peer tutorial sessions	• Total net increase (decrease) in examination performance • Total net increase (decrease) in class participation	• Improved class standing of students • Improved grades of students • Improved academic standing of students

• Available students who are willing to participate in the randomized encouragement design.	as classrooms or conference rooms and the like. • Having dissemination of incentives or nudges for treatment groups (similar to lottery) • Conducting peer tutorial sessions	• Hours of total peer tutorial sessions	• Total net increase (decrease) in project performance • Improved attendance in peer tutorial sessions	• Increased productivity and critical thinking

Key Outcome Indicators

Peer tutorial programs and initiatives have been created with an effort to be able to bolster academic performance of all students choosing to partake in each program. As with any program, there are indicators to performance that are critical in evaluation and must be considered for when evaluating the effectivity of a particular program. Below is a summary of the key outcome indicators found in Table 4.

Table 4 Key Outcome Indicators

Inputs	Activities	Outputs	Intermediate Outcomes	Final Outcomes
• Total Budget in local currency of the peer tutorial program that will be implemented • Total expenses concerned with the peer tutorial program local currency. • Total expenses for the hiring peer tutors • Total number of students available for the	• Approval ratings and satisfaction scores for the peer tutorial students. • Total number of students participating in the peer tutorial sessions • Training expenses for peer tutor	• Total number of peer tutorial sessions conducted • Total number of students who attended the peer tutorial sessions	• Test scores of subsequent exams in a particular course taught in a peer tutorial program • Project and participation scores of subsequent activities in courses taught in a peer tutorial program	• Academic standings of students who participated in the peer tutorial programs • Academic honors of students who participated in peer tutorial programs • Employment and Employment satisfaction of students who

peer tutorial programs • Total number of tutors available for peer tutorial programs	training in local currency • Logistical expenses (venues etc.) in local currency.			participated in peer tutorial programs

Evaluation Design

This section gives and overview of the design of the program evaluation being conducted and how the data shall be analyzed in the models used by the study. It will also highlight impracticalities and ethical considerations that have led to the current observational nature of how the study has been designed.

Impracticality of the Randomized Control Trial (RCT)

The Randomized Control Trial (RCT) is the gold standard of impact evaluation especially when it comes to figuring out what the treatment effect of the program or intervention being studied is (Fowlie and Wolfram, 2010). With this methodology, the sample is drawn out randomly from the population wherein a group is present for both treatment and comparison differentiated by the administering of a treatment and the lack thereof[9]. These two groups have similar characteristics to limit unobserved heterogeneity that may be present in the sample.

Hence, in a randomized control trial, the Average Treatment Effect or the ATE is merely the difference between the impact on treatment and comparison given that full randomization and full compliance were achieved. When there is full compliance, there are no confounders that are present in the study given that all participants had followed their protocol. Furthermore, if there had been full randomization, it can be deduced that the sample chosen (for both treatment and comparison) is representative of the population. This translates into the lack of selection bias in the evaluation. Technically speaking, the selection bias would not be present given these two assumptions in addition to the stable unit treatment value assumption (SUTVA) which states the lack of spillover effects between the two sample groups[10]. Empirically speaking, this deduction is given as

$$E[Y_i|X_i = 1] - E[Y_i|X_i = 0] = E[Y_{1i}|X_i = 1] - E[Y_{0i}|X_i = 1]$$

Observed Difference = Average Treatment Effect on the Treated (ATET)

---

[9] This is typically done through the use of power calculations
[10] There are no general equilibrium effects. Moreover, the treatment and the comparison groups have little to no communication and are directly comparable

$$E[Y_i|X_i = 1] - E[Y_i|X_i = 0] = E[Y_{1i} - Y_{0i}]$$

Observed Difference = Average Treatment Effect (ATE)

However, the RCT may not be used in this particular study given that the mandated participation of a certain proportion of the chosen sample may in turn prevent, hinder or influence the participation of another sample group[11]. This implies the following violations on the assumptions of the RCT. Firstly, the Stable Unit Treatment Value assumption (SUTVA) is violated given that there are spillover effects between treatment and control[12]. Secondly, there is no more full compliance given that while a sample may be representative of the population, there is no direct hindrance for the comparison group to receive the treatment.

This means that the RCT may not be used because the Average Treatment Effect may not serve as an accurate or statistically acceptable basis for the impact of the intervention given that both internal validity and external validity are compromised in the study.

Ethical Considerations

This study also accounts for ethical considerations that must be made with regards to the tutorial program that will be the subject of evaluation. As this shall be purely observational in nature, we will not merely separate or place a distinction on treatment and comparison. We cannot force people who are by all means entitled to attend the program to not attend just for the sake of the study do to ethical and moral implications. Hence, the best treatment effect that may be observed would be purely local in nature. There will only be a separation or distinction purely as a directive from the office of the vice dean if they were recommended to attend the program in the Two-stage least squares regression. There shall be no distinction, however, when the propensity score matching has been conducted.

Description of Population

For this particular study, all respondents will be coming from De La Salle University. In particular, the study will center around students who are taking the course Mathematical Economics II. The total population enrolled in the course Mathematical Economics II are 93 students in the sample. From the 93-eligible population, the proportion between those who had received or engaged in the intervention (which is the tutorial aptitude program) is pegged at roughly 45% of the eligible population.

---

[11] The RCT would not be the best road especially when there is a high chance of imperfect compliance. Hence, an Average Treatment Effect would no longer be obtainable if that would be the case.
[12] SUTVA's assumption on no general equilibrium effects would no longer hold for the most part.

The whole population of students taking Mathematical Economics II are to be used for the paper. To put it simply, the study is centered on a sole subject for baseline measurement as the impact of the tutorial may wildly vary given different subject matter with varying concepts and topics which may lead to misleading inferences.

Process Details

As purely an observational study, the various data points would be collected throughout the different assessments that the students may face. Moreover, various indicators such as grades or certain scores would be obtained from the sole faculty handling the said courses. The table below shows the process that shall be undertaken for the duration of the study.

Table 5 Process Details

Phase 1
Collection of the Data: After a formal assessment for academic indicators has been done (e.g. long exams, projects, final exams, summative assessments, etc.), data from the respective professor/s will be collected and recorded per student together with their attendance to the tutorials.

Phase 2
Regressing the Model: After data had been collected, it may now be regressed using the instrument variable regressions with various covariates used beforehand.

Phase 3
Assessing the Evaluation: The model and the results will pass through a gauntlet of tests such as endogeneity of instruments, exogeneity of regressors, relevance of the instruments, multicollinearity, heteroscedasticity and omitted variable bias.

Phase 4
Forming Policy Recommendations: After the program had been implemented, the results may indicate in favor or opposed to the current structure of peer tutorial programs in the context of DLSU. As only a local average treatment effect (LATE) may be obtained, this result may only be true for the context of DLSU however it may share local effects with other universities of the same characteristics

Instrumental Variable Framework

As an RCT is impractical in the context of this study, a simple regression of the impact on the treatment group and the impact on the comparison group may not be conducted because of the biasedness of the estimates due to the lack of full compliance (a separation) between treatment and comparison (McDuffie et al, 2009).

The following is the specific theoretical underpinning of the Instrumental Variable Regression that will be conducted for the study. The specific regression model will be a Two Stage Least Squares Regression (2SLS)

$$Stage\ 1{:}\ \widehat{X}_i = \alpha_1 + \pi Z_i + u_i$$
$$Stage\ 2{:}\ Y_i = \alpha_2 + \eta \widehat{X}_i + v_i$$
$$Reduced{:}\ Y_i = \alpha_3 + \rho Z_i + e_i$$

The first equation corresponds to the first stage in which we estimate $\widehat{X}_i$ given an instrument. Afterwards, the $\widehat{X}_i$ obtained in the first stage is then used in the second stage to estimate the impact. In the first stage, π is the compliance rate which measures the appropriateness of an observation i in their initial assignment Z_i. Note that eta or η is referred to as the LATE or the local average treatment effect which is not the same as the ATE or the ATET. This is used to estimate the impact on compliers. The LATE can be computed by the formula

$$\eta = \frac{\rho}{\pi}$$

Alternatively, given that an Instrument Variable framework is being used, it may also be adept to utilize the reduced form of the IV. Note that Z_i represents the initial assignment (whether treatment or comparison). ρ would then represent the Intention to treat effect (ITTE) which is unbiased because the treatment assignment is random given that we do not have full compliance.

The first stage of the regression would be the participation hurdle in which attendance in the TAP of the individual is regressed against various determinants. This corresponds to attendance which is a quantitative variable is some function or has some association with the initial assignment that was given to a respondent or the student. This particular relationship is stated econometrically as

$$Attendance_i = \alpha_1 + \pi D_i + \beta_1 ProblemSet_i + \beta_2 Sex_i + \beta_3 MATECO1Grade_i + U_i$$

Table 6 shows the descriptions of the variables found in the first stage of the two stage least squares regression. These variables were similar to both the theoretical underpinning of the 2SLS model for impact evaluation as well as the study conducted by Paloyo et al. in 2016. While this is an observational study, we base treatment assignment on the simple rationale if they had attended the tutorial program or had engaged in services in accordance with this.

Table 6 Two Stage Least Squares Stage 1 Variables

Name	Description
Attendance	Attendance is a continuous positive or zero value which indicates the number of sessions attended by the respondent. This is the dependent variable of the first stage of the 2SLS regression
D_i	The participation indicator is a binary variable which corresponds to "1" if the respondent was assigned to the treatment group (attended TAP) and "0" if the respondent was assigned to the comparison group (has not attended TAP). This is the independent variable of the first stage of the 2SLS regression.
Problem Set	Problem Set is a continuous positive or zero value which indicates the score of the student in the problem set prior there being any TAP sessions for MATECO2. This is an independent variable of the first stage of the 2SLS regression.
Sex	Sex is a binary variable corresponding to "1" if the student is male and "0" if the student is female. This is an independent variable of the first stage of the 2SLS regression.
MATECO1 Grade	MATECO1 Grade is a continuous positive value which indicates the grade received by the students (numerical equivalent) in the prerequisite course of MATECO1. This is an independent variable in the first stage of the 2SLS regression.

The second stage of the regression would correspond to the impact hurdle which corresponds to determining the impact of the tutorial program as an effect on a main academic indicator. For this study, per the objectives, the evaluation centers on the ability of the peer tutorial program to affect the academic performance of a student. In the context of this study, academic performance may be indicated or represented by the final grade of that student in a particular subject. This is regressed against the attendance or the number of times a student has attended the peer tutorial sessions and various covariates. Econometrically speaking, this relationship or association is given as

$$FinalGrade_i = \alpha_2 + \eta Attendance_i + \beta_1 Midterm_i + \beta_2 Finals_i + \beta_3 ClassParticipation_i + \beta_4 Sex_i + v_i$$

Table 7 shows the descriptions of the variables found in the second stage of the two stage least squares regression. These variables were similar to both the theoretical underpinning of the 2SLS model for impact evaluation as well as the study conducted by Paloyo et al. in 2016.

Table 7 Two Stage Least Squares Stage Two Variables

Name	Description
Final Grade	Grade is a continuous quantitative variable wherein it measures the total grade that has been received by the student in the course. This is not the categorical measurement, rather, the numerical equivalent of the total score threshold a student was able to obtain. Grade is the base dependent variable of the 2SLS regression and the main dependent variable of the second stage of the 2SLS regression.
Attendance	Attendance is a continuous quantitative variable which measures the total number of times a student had attended peer tutorial services. The coefficient from this variable in the second stage of the 2SLS regression would be the Local Average Treatment Effect or the LATE. This is an independent variable in the second stage of the 2SLS regression.
Midterm	Midterm is a continuous quantitative variable which measures the score that a student was able to obtain in an assessment conducted by the professor for the midterm period. This is an independent variable in the second stage of the 2SLS regression.
Finals	Finals is a continuous quantitative variable which measures the score that a student was able to obtain in an assessment conducted by the professor for the final examination period. This is an independent variable in the second stage of the 2SLS regression.
Sex	Sex is a binary variable which corresponds to "1" if the respondent is male and "0" if the respondent is female. This is an independent variable in the second stage of the 2SLS regression.

Propensity Score Matching Framework

Apart from the typical framework of an Instrumental Variable Regression, we may also use the framework of Propensity Score Matching. This methodology is an impact evaluation methodology for observational studies wherein the determination of treatment and comparison were not random (Gertler et. al, 2013). This involves calculating a propensity score in which the score represents the likelihood by which a unit observation may be segmented into the treatment group based on pre-treatment characteristics. After which, provided that there is common support, the treatment effects shall be estimated. For this case, the matching methodology shall be the nearest neighbor approach, the radius approach and the kernel approach.

Calculating the propensity score would need the use of the Probit regression model. In this regression model, we will determine the likelihood, or more specifically, the propensity of observations to be assigned into treatment groups. We specify first this participation or

treatment variable as a variable D_i where this variable denotes the subjection of an individual i to either have participated or not participated in the treatment which in this case is the tutorial aptitude program.

$$D_i = \begin{cases} 1 & if \ \ TAP \ Attended \\ 0 & if \ TAP \ Non \ Attended \end{cases}$$

Hence, the population may be subjected into two sub classifications wherein one group received the treatment (or the TAP) while the other group were not able to receive the treatment (i.e. had not attended the tutorial aptitude program).

$$N_{PSM} = n_{(D_i=1)} + n_{(D_i=0)}$$

As this is an observational study, there would definitely be unbalanced numbers in the division between our treatment and comparison groups given that there is no control over the said distinction.

We calculate this propensity score $\widehat{P_i}$ using the Probit Regression Model below. Note that we will only be using pre-treatment characteristics in estimating the propensity score.

$$\widehat{P_i} = \Pr[D = 1 | X_i] = E[D|X]$$

$$D_i = \alpha_0 + \alpha_1 ProblemSetScore_i + \alpha_2 Sex + \alpha_3 MATECO1Grade_i + \epsilon_i$$

Table 8 Propensity Score Determination Variables

Name	Description
Problem Set Score	Attendance is a continuous positive or zero value which indicates the number of sessions attended by the respondent. This is an independent variable in the propensity score determination.
MATECO1 Grade	MATECO1 Grade is a continuous positive value which indicates the grade received by the students (numerical equivalent) in the prerequisite course of MATECO1. This is an independent variable in the propensity score determination.
Sex	Sex is a binary variable which corresponds to "1" if the respondent is male and "0" if the respondent is female. This is an independent variable in the propensity score determination.

For the calculation of the Average Treatment Effect and the Average Treatment Effect on the Treated, we shall be using the determinants given in the second stage of the two-stage least squares regression as our main determinants.

Data Collection Plan

       This section delves into the data points and how these data will be used in the context of the quasi experimental techniques, namely, two-stage least squares and propensity score matching. The formal analysis plan also highlights how the methodology shall be carried out and the key estimates that may be obtained from these models.

Data

       Academic indicators and the attendance to the peer tutorial programs are the main data points that are needed in order to conduct disaggregations and regressions to be able to estimate the Local Average Treatment Effect (LATE) from the Two Stage Least Squares Framework. Data would need to represent the key outcome indicators that is expected for each stage of the study in order to determine the impact of peer tutorial programs on the academic performance of students. Data needed for this study are highlighted in Table 8 wherein the specific data point and contribution to the study are laid out.

Table 9 Data Points

Specific Data Point	Description of the Data Point	Contribution to the Study
Final Grades	Final Grades correspond to the grade point (in percentage unit) a student was able to obtain for his/her performance in a subject for a given term. This is reported at the end of the term and may be obtained from the professor of the subject under evaluation for the study	Final Grades will make up the Grade values in the second stage of the 2SLS regression model and will serve as the base academic indicator in the framework to be utilized.
Test Scores	Test scores correspond to the average or the aggregate scores a student was able to obtain from a recorded examination or performance assessment in a particular subject.	Test Scores will make up the Test Score values in the second stage of the 2SLS regression model and will serve as a covariate to explain grade.
Attendance of Student (Class)	Attendance of a Student is the number of times a student has attended a course throughout the term. This excludes all absences for a particular session in a course and is measured in the number of times.	Attendance will make up the Class Attendance values in the second stage of the 2SLS regression model and will serve as a covariate to explain grade.

Attendance of Student (Attendance)	Attendance of a Student is the number of times a student has attended peer tutorial sessions throughout the duration of the experiment. Attendance wherein the experiment was not conducted shall be excluded from the count.	Attendance will make up the Attendance values in the first stage of the 2SLS regression model.

Analysis Plan

The Two stage least squares regression will provide the econometric framework by which the local average treatment effect (LATE) may be obtained as the coefficient of the second stage of the regression. The LATE, as the name implies, is a local treatment effect which means that the value only holds true for the sample that the study is currently observing. Once instrumental variables and two stage regressions are used, biasedness is assumed in the model (Paloyo et al., 2016). Given that there is a presence of sample selection bias due to the selection of De La Salle University students taking Mathematical Economics, it is needed that a model be used that can account for this biasedness or consider its presence.

The formal analysis plan involves regression of the model, interpretation of the model, formal testing and recommendation based on the results. This process is detailed in table 10.

Table 10 Two Stage Least Squares Analysis Plan

Step	Description
Regression	A Two Stage Least Squares regression will be conducted based on the specification given as $$Attendance_i = \alpha_1 + \pi D_i + \beta_1 ProblemSet_i + \beta_2 Sex_i + \beta_3 MATECO1Grade_i + U_i$$ $$FinalGrade_i = \alpha_2 + \eta Attendance_i + \beta_1 Midterm_i + \beta_2 Finals_i + \beta_3 ClassParticipation_i + \beta_4 Sex_i + v_i$$ The p-values, z values and corresponding coefficients would be reported with the value of η being of particular importance as it will correspond to the Local Average Treatment Effect.
Interpretation	Each coefficient will be interpreted in the context of how it is able to explain the variation in the grades that a student receives at the end of the term. All pertinent information will be listed including significance at the average impact of the peer tutorials on the grade of the student.
Formal Testing	The model shall be subjected to tests for instrument relevance and instrument exogeneity.
Recommendation	Through the LATE, a recommendation is to be made on future action with regards to the implementation and continuity of peer tutorial programs in the context of DLSU Economics majors program.

While the two stage least squares regression may only yield a LATE, the Propensity Score Matching technique may yield both an Average Treatment Effect (ATE) and an Average Treatment Effect on the Treated (ATET). The formal analysis plan for the Propensity Score Matching Framework may be seen below.

Table 11 Propensity Score Matching Analysis Plan

Step	Description
Identifying the Propensity Score	A Probit regression is to be done wherein we determine the propensity score against pre-treatment variables specified in the formulation below.
Identifying Common Support	Based on the propensity score calculated, the common support region between the two segmentations shall be determined.
Identifying the Optimal Blocks	After the propensity score has been identified, the optimal number of blocks shall be determined prior to the matching stage.
Satisfying the Balancing Property	Once the blocks have been determined, the propensity score shall adjust the given treatment and comparison (i.e. went to TAP and did not go to TAP) and determine whether the division is balanced.
Estimating ATE and ATET	The ATE and the ATET shall be estimated using Nearest Neighbor Matching, Radius Matching and Kernel Matching methods based on the identified optimal blocks.
Policy Recommendations	Through the ATE and ATET, a recommendation is to be made on future action with regards to the implementation and continuity of peer tutorial programs in the context of DLSU Economics majors program.

Results and Discussion

In this section, we report the results of the two-stage least squares regression and the propensity score matching estimation that was conducted. We first give an overview of the variables using various descriptive statistics. We also give meaning to the regression and matching results from the estimated models and seek to answer the main objective of the study.

Descriptive Statistics

Below we see some general descriptive statistics of the variables in the study generated using Gretl. Various measures have been shown in order to give a better picture of the data set that was used.

Table 12 General Descriptive Statistics

Variable	Mean	Median	Minimum	Maximum
Problem Set	32.962	38.000	0.0000	45.000
Midterm Exam	31.167	31.500	13.500	39.500
Final Exam	45.796	49.000	0.0000	68.000
Final Grade	80.314	82.295	16.810	101.81
MATECO1 Grade	82.516	86.000	53.000	100.00
Variable	**Std. Dev.**	**C.V.**	**Skewness**	**Ex. kurtosis**
Problem Set	10.614	0.32200	-1.2967	1.0239
Midterm Exam	5.9629	0.19132	-0.46016	-0.53448
Final Exam	13.701	0.29918	-0.91595	0.69973
Final Grade	15.077	0.18773	-1.1604	2.2992
MATECO1 Grade	13.076	0.15847	-0.77733	-0.44656
Variable	**5% Perc.**	**95% Perc.**	**IQ range**	**Missing obs.**
Problem Set	11.000	44.000	14.000	0
Midterm Exam	20.500	39.000	10.250	0
Final Exam	22.200	63.800	22.000	0
Final Grade	54.275	98.905	21.546	0
MATECO1 Grade	55.400	98.300	20.000	0

Comparing Treatment and Non-Treatment Groups

In order to see a clearer picture between the group which had not received the treatment (i.e. had attended a TAP Class) and the group which had received the treatment (i.e. had attended a TAP Class), we generate some descriptive statistics on this matter. Below are figures that elucidate differences or similarities between the two groups on some key variables.

Table 13 Initial Comparison between Treatment and Non-Treatment

Classification	Treatment Group (Did Attend TAP)	Non-Treatment Group (Did Not Attend TAP)
Total	51	42
Percentage of Total	0.5484	0.4516

In the table above, we can see that there seems to be a relative balance between treatment and non-treatment which is essential in studies involving matching. An adjusted comparison will be seen later when we have adjusted these for the propensity score.

Below we see a more detailed figure that highlights differences and similarities between treatment and non-treatment in key outcome indicators.
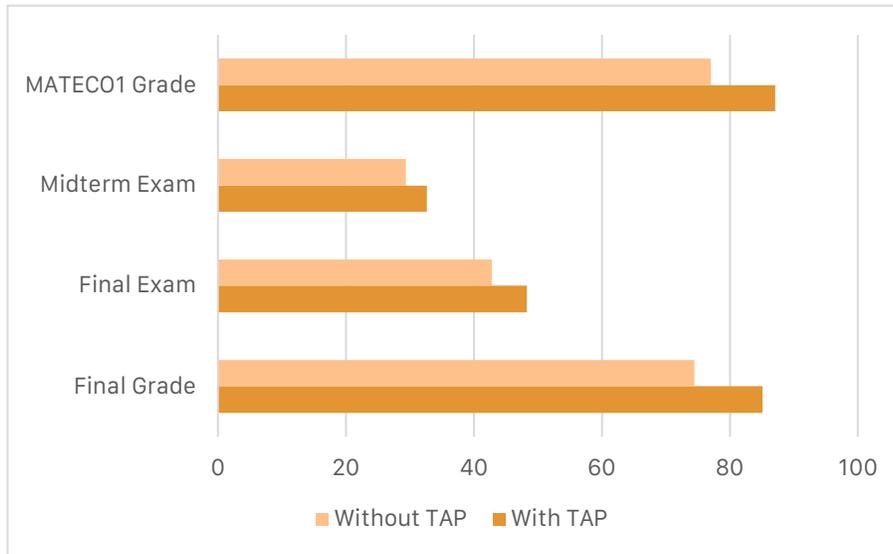
Figure 1: Key Outcome Indicators of Treatment and Non-Treatment TAP Groups

Notice that to begin with, the MATECO1 grade of the students who attend TAP are somewhat higher compared to those who do not attend TAP. The same trends appear with other outcome indicators such as the scores in the midterm exam, the final exam and subsequently the final grade. In here, we can see that majority of students engaging in the tutorial program generally have higher grades to begin with, however, the difference is slightly low making the two groups generally comparable (see balancing in the propensity score matching results).

Two Stage Least Squares Regression Results

The Instrumental Variable regression results are given in the table below following the model specification discussed previously. The instrumental variables used were the participation dummy, the scores obtained in a problem set prior to the commencement of the TAP, the grade of the student in a requisite course (MATECO1) and the gender of the student. Below are the results of the two-stage least squares regression.

Table 14: Two Stage Least Squares Regression Results

	Dependent Variable: Final Grade in MATECO2	
Variable Name	**Coefficient (Standard Error)**	**z-statistic**
Attendance	2.26434*** (.6501828)	3.48
Final Exam	.6360097*** (.0709688)	8.96

21

Midterm Exam	.7285841*** (.1675534)	4.35
Class Participation	.0435946 (.0541944)	0.80
Model Intercept	24.25175*** (3.736802)	6.49

*,**,*** indicate significance at a 90%, 95% and a 99% Confidence bound respectively

As seen in this regression, the final grade is very much associated with the final exam scores a student has obtained, the scores of the midterm examination and the attendance of the student in the tutorial aptitude program. Note that these variables are all positively correlated indicated direct gains in the final grade as well as being significant at a 99% confidence bound. In particular to this study, notice the coefficient of attendance which corresponds to a 2.26434-point increase in the final grade. This seems to suggest that an additional tutorial aptitude program session may correspondingly increase the grade of the student by 2 basis points which is a relatively big magnitude considering a scale only until 100.

As the coefficient of the variable Attendance is the Local Average Treatment Effect, we can see that the LATE is significant in the estimation of the dependent variable (our outcome variable). Moreover, it positively with the said outcome variable suggesting that the treatment which is the tutorial aptitude program had positively impacted the performance of a student at least indicated by the final grade. We specify the details of the Local Average Treatment Effect from the Two Stage Least Squares Regression below.

Table 15 Local Average Treatment Effect

LATE	Z	P > |z|	Standard Error	F-Stat p-value (Model Significance)	R-squared
+ 2.264634	3.48	0.0000	.6501828	0.00000	0.8330

Therefore, based on the two-stage least squares regression in estimating the Local Average treatment effect, we can conclude that attendance in the Tutorial Aptitude Program (TAP) was indeed able to increase the final grades of the students which is a key outcome indicator of the program. Post estimation testing for the Exogeneity of the instruments which are the Durbin (1954) and the Wu (1974) and Hausman (1978) indicate that the instruments are indeed exogenous. Moreover, Shea's Partial R^2 and the First stage tests for the validity of the instruments reject the null hypothesis that the instruments are weak. We show these tests below.

Table 16 Two Stage Least Squares Post Estimation

Test	Statistic and Conclusion

Durbin (score)	1.4145 (0.2343) Instruments are Exogenous
Wu-Hausman	1.34367 (0.2496) Instrument are Exogenous
First Stage Test	0.7831 (0.0000) Instruments are Valid

As can be seen in the results, it seems as though the instruments were valid and exogenous suggesting that there is no problem of weak instruments that may be present in the model.

Propensity Score Matching Results

The Propensity Score Matching methodology suggested significant gains in the final grade of the students as reported by the Average Treatment Effect and the Average Treatment Effect on the Treated. We will first run down the different methods before reaching the desired treatment effects.

Calculating the Propensity Score

We calculate the propensity score using the Probit model we had specified earlier. This estimation consists of pre-treatment variables which calculates the propensity score or the probability that a person would select or be part of a treatment group. The results of the probit regression are seen below.

Table 17 Calculating the Propensity Score Using Probit

	Dependent Variable: Participation Dummy Variable	
Variable Name	**Coefficient (Standard Error)**	**z-statistic**
Problem Set	–0.0153135 (.0181807)	–0.84
Sex	–.4639154* (.2876928)	–1.61
MATECO1 Grade	0.0455885** (.0150206)	3.04
Model Intercept	–2.911729** (1.02218)	–2.85

*,**,*** indicate significance at a 90%, 95% and a 99% Confidence bound respectively

As the propensity score was calculated, a region of common support which is the region in which matching may take place may be identified. A breakdown of the estimated propensity score in the region of the common support region may be seen below.

Let's Do It Right!

Central to the Propensity score matching methodology is the feasibility of the matching when there is common support and when balancing has been properly satisfied. While matching may be conducted without these, adequate estimation may only be achieved when we have accounted for these factors. We now do it right by accounting for common support in the table below.

Table 18 Propensity Score Estimates Accounting for Common Support

Estimated Propensity Score on Common Support [.13572204, .83866524]

Level	Bounds	
1%	.135722	.135722
5%	.1683101	.136183
10%	.2040424	.1536958
25%	.4488844	.1677794
50%	.6176223	
75%	.6955137	.8272614
90%	.7857239	.8311481
95%	.8324096	.8385791
99%	.8386652	.8386652

As the propensity scores have now been calculated, we need to verify whether the balancing property has been satisfied. We show the blocks and the balancing to show that the balancing property had indeed been satisfied.

Table 19 Balancing of Participation in TAP

Inferior of the Block Propensity Score	Has Not Attended the TAP	Has Attended the TAP	Total
.135722	5	3	8
.2	12	1	13
.4	8	14	22
.6	14	28	42
.8	1	5	6
Total	40	51	91
D_i is a balanced ratio in the area of common support			

As seen in the table above, there is indeed a satisfactory balancing between those that had attended the tutorial program and those that have not even adjusting for the propensity score and the optimal number of blocks.

Calculating Treatment Effects

As the participation is balanced, we can now commence with the matching methodology to calculate for the treatment effects. Below are the treatment effects using the Nearest Neighbor, Kernel and Radius Matching Methods.

Table 20 Average Treatment Effect and Average Treatment Effect on the Treated

Matching Method	Average Treatment Effect (ATE)	Average Treatment Effect on the Treated (ATET)
Nearest Neighbor Matching	4.136174**	4.54319*
	(1.82)	(1.52)
Radius Matching	2.621	2.674***
	(0.751)	(3.109)
Kernel Matching	2.492	2.579
	(-)	(0.851)

*,**,*** indicate significance at a 90%, 95% and a 99% Confidence bound respectively

As we can see, some of the treatment effects were significant while the others were not. Note that the standard errors in the Radius Matching and the Kernel matching were bootstrapped in order to be more robust.

We can see that the Average Treatment Effects from the nearest neighbor matching indicate that there are significant gains from the program which leads to a roughly 4-point increase in the final grade. This is once again a significant increase like those reported by the LATE in the conducted 2SLS regression. However, the magnitude is almost double that of the LATE likely because of near matches using the nearest neighbor method. The radius matching method reported a significant ATET but an insignificant ATE which had matched the results of the LATE in the 2SLS regression approximately. The kernel ATE and ATET were found to be insignificant likely because the sample is extremely small. With these results, we can see that while significant effects were obtained, a small sample size like the one in this study has led to the insignificance of some treatment effects in addition to a failure to randomize (as there is no distinction between treatment and control) as well as no delineation between treatment and comparison.

Despite these limitations, we can see that the tutorial aptitude program was able to induce positive effects to a key outcome indicator once again suggesting that the program was able to meet its objectives and goals.

Policy Recommendations

As we have seen in the results, the TAP seems to be an effective program in eliciting higher grades from students that have participated in the program. Hence, it may be imperative to increase the reach of this activity to even more subjects and even more settings outside of the economics program. While we did calculate for mere local effects, it may be an easy transfer to other colleges given similar characteristics of students and subjects.

Expanding Reach

The economics department and indeed the school of economics is no stranger to innovating best practices and proper policies backed up by sound empirical evidence and data. As such, the best practice of the economics department in peer tutorial programs may be expanded by allowing a better recruitment system for tutors given that the program has been validated to have positive associations with key outcome indicators that need to be satisfied by peer tutorial programs.

Inclusivity Measures by the Associate Dean

As the program has been seen to induce positive effects, the Associate dean may recommend more students who are in dire need to help to raise their academic standing or have had a substantial amount of accumulation to target these students to participate in the program. That way, there will be a lower attrition rate in the school of economics and other schools which are notorious with stringent cutoffs and high drop-out rates.

Summary and Conclusions

Peer tutorial programs have been rampant in their reach throughout many universities and colleges not just in the Philippines, but the world in general (West et al., 2008). However, the effectivity of such programs has not been placed in impact evaluation rigors especially in the Philippine context. The study seeks to determine the effectivity of peer tutorial programs with end term grades as the main outcome variable. The local average treatment effect has been estimated using the Two Stage Least Squares regression using the first stage and the second stage method vis-à-vis the reduced form estimation and was significant. The average treatment effect and the average treatment effect has been generated from the propensity score matching and had mixed but generally positively significant results. The insignificance of some treatment effects may have been merely due to the low sample size for the study as matching methods typically demand very large data sets in order to be effective. Overall, we have seen that there is a positive effect on the key outcome indicators for the students who have partaken in this program suggesting that it has met some of its key goals.

References

Burns, E. (2006). Pause, prompt and praise - peer tutored for pupils with learning difficulties. *British Journal of Special Education*, *33*(2), 62-67.

Fowlie, M., Wolfram, C. (2010). *An experimental evaluation of the federal weatherization assistance program*.

Harper, G., & Maheady, L. (2007). Peer-mediated teaching and students with learning disabilities. *Intervention in School & Clinic*, *43*(2), 101-107.

McDuffie, K., Mastropieri, M., & Scruggs, T. (2009). Differential effects of peer tutoring in co-taught and non-co-taught classes: results for content learning and student–teacher interactions. *Exceptional Children*, *75*(4), 493-510

Paloyo, A., Rogan, S., Siminsky, P. (2016). The effect of supplemental instruction on academic performance: An encouragement design experiment. *Economics of Education Review*. 55:57-69

West, S. G., Duan, N., Pequegnat, W., Gaist, P., Des Jarlais, D. C., Holtgrave, D., … Mullen, P. D. (2008). Alternatives to the Randomized Controlled Trial. *American Journal of Public Health*, *98*(8), 1359–1366. http://doi.org/10.2105/AJPH.2007.124446