Machine Learning based Analysis of the Effect of Team Competition on College Students’ Academic Performance

Huiling Zhang  
Shandong Vocational and Technical University of International Studies

Huatao Wu  
University of Massachusetts Amherst

Zhengde Li  
Qufu Normal University

Wenwen Gong  
Tsinghua University

Yan Yan (✉️ yanyan@qdu.edu.cn)  
University of Qingdao

Research Article

Keywords: Team Competition, Prediction, Individual Treatment Effect, Field Experiment

Posted Date: November 1st, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3519859/v1

License: ☑️ This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Journal of Cloud Computing on February 9th, 2024. See the published version at https://doi.org/10.1186/s13677-024-00591-2.
Machine Learning based Analysis of the Effect of Team Competition on College Students’ Academic Performance

Huiling Zhang¹, Huatao Wu², Zhengde Li³, Wenwen Gong⁴, Yan Yan⁵, *

¹ School of Foreign Languages, Shandong Vocational and Technical University of International Studies, Rizhao, 276826, China
² College of Engineering Electrical & Computer Engineering, University of Massachusetts Amherst, United States
³ School of Computer Science, Qufu Normal University, Rizhao, 276826, China
⁴ Department of Computer Science and Technology, Tsinghua University, China
⁵ College of Quality and Standardization, University of Qingdao, Qingdao, 266071, China
*Corresponding Author: Yan Yan. Email: yanyan@qdu.edu.cn

Abstract: In the field of Mobile Edge Computing (MEC), machine learning techniques present a promising avenue for intelligent integration and processing of data from MEC terminals. Our study delves into the intersection of Machine Learning with MEC terminal data, exploring the complexity of team competition mechanisms based on social identity and competition theory. This exploration aims to enhance student participation and enthusiasm within university classrooms. However, despite of its potential benefit, there are still many unresolved issues: What type of students and teams benefit more from team competition? In what teaching context is team competition more effective? Which competition design methods better increase student academic performance? To answer these questions, we first design a randomized field experiment among freshmen enrolled in college English course. Then, we collected data using mobile devices and analyzed the observational data to predict the individual treatment effect of academic performance in college English through linear and nonlinear machine learning models. Finally, by carefully investigating features of teams and individual student, we reduce the prediction error by up to 30%. In addition, through interpreting the predictive models, we discover some valuable insights regarding the practice of team competition in college classrooms.

Keywords: Team Competition; Prediction; Individual Treatment Effect; Field Experiment

1. Introduction

In recent years, with the widespread proliferation of smart terminal devices, Mobile Edge Computing (MEC) applications have aggregated substantial data at the terminal level [1], emerging as a potent force reshaping the dynamics of classrooms. By bringing computational capabilities closer to end-users, MEC facilitates the seamless integration of digital resources within the educational domain [2]. Within this context, MEC terminals act as intelligent hubs which capture valuable teaching data in real-time. Team competition strategy [3] [4], serving as motivational mechanisms, has found extensive adoption in various educational tiers. This strategy, is not limited to classroom settings but has been further developed, particularly with the support of MEC. It fosters positive competition among students, enhancing learning efficiency. Additionally, it provides teachers with greater data support, helping them better understand students’ learning needs. Hence, it plays a significant role in modern education.

Despite the potential benefit of team competition, plenty of unknowns remain. Because of the huge heterogeneity among the schools, the majors, the classes and the students, which may lead to significant variations in students’ motivation and academic performances? What types of students and teams (i.e., gender, major, grade) benefit more from team competition? Which teaching design methods (i.e., team formation) better increase students’ academic performance? In what teaching context team competition is more effective? Whether there is a causality between in-class activities (i.e., discussion, quiz, homework, etc.) and academic performance. Understanding the causal effects between these factors and students’ academic performance can help teachers optimize the practice of team competitions in college classrooms for different types of students, thereby improving students’ motivation and academic performance.

However, it is challenging to answering these questions. First, there are few real-world data which covers the whole team competition learning process. Controlled field experiments are necessary to collect enough data for our
research. Second, measuring the causal effects between the team competition mechanism and students’ academic performance is intrinsically difficult\cite{5}. It requires a proper definition of individual performance measures and prediction targets\cite{6}. Third, the variable space to describe the characteristics of context, students, team and teaching activities is high-dimensional\cite{7-9}. Moreover, there are a lot of complex relationships among them. Domain knowledge and data analytics are both needed to identify the potential predictive factors\cite{10}.

In this paper, we propose a novel approach to attack these challenges, as shown in Figure 1. We first develop a randomized field test among freshmen enrolled in college English course, then we predicting the individual treatment effect of team completion on students’ academic performance. Moreover, through interpreting the predictive models, we investigate the most significant factors in the practice of team completion in college classrooms. Since students’ performance in teaching activities i.e., answer race, discussion, quiz and homework are distributed over long periods of a semester, we need to integrate their homework results data into a central cloud platform called online teaching platform, for more comprehensive data analysis and mining.

Concretely, our contributions include:

1. We design and conduct a controlled field experiment to collect data covering the whole process of team competition learning.
2. We formulate the problem as a prediction task. And we adopt machine learning models to predict students’ individual treatment effect of team competitions
3. We interpret the prediction model to identify the most important factors in team competition learning.

2. Related Work

Team competition. As an incentive mechanism based on social identity and contest theories, team competitions have been increasingly applied in many fields. It has shown that team competition can not only effectively improve key metrics, i.e., participation\cite{11}, but also help them obtain a sense of achievements \cite{16}. Markus et al. \cite{14} investigate how to leverage team competition to improve the cost efficiency in crowdsourcing through a large-scale experimental evaluation. Ai et al. \cite{15} conduct an inter-team contest field experiment on a ride-sharing platform, and find that drivers participated in the team competition works longer hours and earn higher revenue than drivers in control conditions. Ye et al. \cite{12} study how different factors of team completion affect the outcomes of individual drivers in ridesharing based on the result of the online field experiments.

With regards to education, the imperative to maintain competitiveness and facilitate the transformation of database management practices has necessitated alignment with the prevailing, cutting-edge technological trends within the industry \cite{16}. DiNapoli \cite{4} describes the implementation of a pedagogy based on team competition in mathematics classrooms. It shows that team competition could be a useful motivator. Scales et al. \cite{11} conclude that team-based game mechanics can increase resident participation in an online learning platform delivering quality improvement content. They draw the conclusion through a randomized, controlled field experiment. To enhance the
effectiveness and quality of experimental teaching, a comprehensive experimental teaching course system that combines artificial intelligence and edge computing technologies is built[17]. By deploying edge computing nodes in laboratories or educational settings, experimental data can be transmitted in real-time to edge devices for processing and analysis. Such as students’ respective health physique data is integrated into a central cloud platform for more comprehensive data analysis and mining[18]. However, to the best of our knowledge, few have analyzed the importance of different characteristics in team competition, particularly in college English teaching.

Individual treatment effect prediction. Predicting individual treatment effects of actions plays a critical role in many domains[20–23]. Synthetic Minority Oversampling TEstchnique (SMOTE) technique is used for pre-processing the missing value in the provided input dataset to enhance the prediction accuracy[23]. A new Metaheuristic Optimization-based Feature Subset Selection with an Optimal Deep Learning model (MOFSS-ODL) for predicting students’ performance is presented[24]. Many researchers propose a variety of algorithms for predicting ITE based on different techniques, i.e., deep neural networks [25], random forests [26], etc. Others study the application of ITE prediction in different fields, i.e., medicine [26–27], online platforms [29]. Our work is similar to recent work that predicts ITE in a ride-sharing economy [12]. However, our work focuses on the ITE prediction of students’ academic performance. Moreover, different models are adopted to better capture the characteristic in college English teaching.

3.1 Experiment Setup

To test the impact of team competition on the academic performance of college students, we develop a randomized field experiments among freshmen enrolled in college English course. We choose college English course to conduct a classroom experiment for two reasons. First, as part of commonly required courses, college English has a large enrollment in Chinese universities. The assessment of this course is highly standardized. All students utilized identical course materials, with instructional activities and examinations administered through a unified online platform hosted on the Mobile Edge Computing (MEC) terminal. Therefore, this course structure allows us to split control and treatment groups among classrooms uniformly. Second, the direct link between students’ academic performance and scholarships, graduation and post-graduation employment provides motivation for students to do well in college English course.

Our sample is made up of freshmen enrolled in college English course taught by the author during the fall semester of the 2021-2022 academic year. We exclude students with incomplete information, resulting in a final sample of four classes and 180 students. Error! Not a valid bookmark self-reference. shows the descriptive statistics for students in different groups. The first row shows the number of observations in each group. The second row demonstrates the ratio of female students in each group. The ratio of students from Shandong province, where the university located, is shown in the third row.

<table>
<thead>
<tr>
<th></th>
<th>Control</th>
<th>Treatment 1</th>
<th>Treatment 2</th>
<th>Treatment 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observations</td>
<td>41</td>
<td>48</td>
<td>49</td>
<td>42</td>
</tr>
<tr>
<td>Female</td>
<td>0.707</td>
<td>0.479</td>
<td>0.306</td>
<td>0.714</td>
</tr>
<tr>
<td>Within-province</td>
<td>1.000</td>
<td>0.958</td>
<td>0.837</td>
<td>0.881</td>
</tr>
</tbody>
</table>

3.2 Team Formation

We randomized classrooms into either a control group or one of three treatment groups, as shown in Error! Reference source not found.. In the first treatment group, students are permitted to create teams freely. In our second treatment group, students are assigned to different teams randomly. This group is intended to replicate the most common scenario of team formation in teaching practice. In the third treatment group, students are split into different teams according to their academic performance, i.e., the score of English in National College Entrance Examination (NCEE). The Control group uses traditional teaching methods, indicating that no team competition mechanism is introduced in teaching process. All the teams shaped in similar size, covering 6 to 7 regular members.
3.3 Contest Design

During the contest period, all teams in three treatment groups will engage in team competitions to compete with other teams in the same class. And scores will be rewarded to these teams according to their ranks in the class. The score will contribute to the final score of the course. Besides final exam, the final score of a student also includes performance in teaching activities, i.e., answer race, discussion, quiz and homework. All the activities are conducted on an online teaching platform, and the performance of students are collected automatically. We denote the score of a team by averaging the final score of all team members. The scores of each team members and other teams are presented on score board for students to check during the contest period. At the end of the semester, top 5 teams on the score board in each treatment group will be rewarded 5 to 10 extra points to their final score.

4. Predicting the Individual Treatment Effect

4.1 Problem Formulation

The individual treatment effect (hereafter referred to as ITE) indicates the effect of team competitions on the academic performance of a student. We employ difference-in-differences (DID) approach [30] to estimate the ITE. The DID approach first calculate the difference in academic performance before and after team competition for each student; average the performance change in control group, and compute the difference between the two conditions.

Formally, given a student set $S = S_{t1} \cup S_{t2} \cup S_{t3} \cup S_c$, where $S_{t1}, S_{t2}, S_{t3}$ and $S_c$ indicate students in treatment group 1, treatment group 2, treatment group 3 and control group, respectively. Let $S_{i,T}$ be the academic performance of student $i$ in the time period $T$, $T_0$ be the baseline period before competition starts, and $T_1$ be the time period when the competition ends. The difference of student $i$ in academic performance before and after competition period can be calculated by

$$\Delta S_i = S_{i,T_1} - S_{i,T_0} \quad (1)$$

And the average performance difference of students in control group can be calculated by

$$\Delta S_{control} = \frac{\sum_{i \in S_c} \Delta S_i}{|S_c|} \quad (2)$$

Finally, the individual treatment effect of student $i$ can be obtained by

$$ITE_i = \Delta S_i - \Delta S_{control} \quad (3)$$

Given a student $i$ in team $j$, let $F_{S_i}$ denote the feature list of the student, and $F_{T_j}$ represent the features of team $j$. The problem of predicting the ITE of student $i$ can be formulated by

$$\hat{ITE}_i = f \left( F_{S_i}, F_{T_j} \right) \quad (4)$$

4.2 Feature Selection

Based on the theoretical insights from social identity theory and contest theory [30–31], as well as the domain knowledge from college English teaching, we characterize the features of a student in our experiment from two aspects: team features and individual student features.
4.2.1 Team Features

Team features depict the team-level characteristics that are related to the behavior of students, such as team formation strategy, team diversity and average performance of a team. In detail, team diversity is indicated by gender diversity and hometown diversity, which are measured by the ratio of female students and students within the province. To depict the performance of a team, we average all the teammates’ Aptis grades. The performance of a team is a potential significant predictor of ITE.

4.2.2 Individual Student Features

Individual student features are made up of the demographics, academic performance before the competition, and classroom behaviors of a student. To depict student academic performance before the competition, we investigate students’ performance in National College Entrance Examination (NCEE) and Aptis test. In detail, NCEE performance is indicated by overall mark and subject marks. Aptis performance is indicated by the overall score, scores of listening, speaking, reading and writing, and a score for the grammar and vocabulary component. We capture students’ classroom behaviors from three aspects: times of participating answer race, scores of quiz and homework. Moreover, student demographics, e.g., gender, hometown and age, are also contained in the set of features.

In our study, a student’s ITE is calculated by its Aptis score and the score of final exam. Aptis is an assessment tool which is widely adopted in China. It can help accurately test English language abilities in all four skills, reading, listening, writing and speaking. It is held in every October in our school to assess the English language level of our students. All the freshmen are asked to participate in the exam, which provide us with a fully and accurate evaluation of students’ English ability before the competition. The distributions of students’ Aptis score in each group are approximately normal, as shown in Figure 3.

![Figure 3. Distributions of Aptis overall marks of all participants and three treatment groups.](image)

Final exam is conducted at the end of the semester, which includes written and oral test. All the groups use the same test paper and mark by the same teacher. Because the result of oral test may be subjective, we only take the score of written test to calculate the ITE of a student. The distribution of final exam scores of all the participants in each groups is demonstrated in Figure 4.

![Figure 4. Distributions of final exam marks of all participants and three treatment groups.](image)

4.3 Model Implementations

A number of machine learning models can be employed for ITE prediction. Because our study focus on understanding the potential predictors for ITE, we only consider models that can easily interpret the importance of all the influential factors. Here we choose four commonly used machine learning methods: extreme gradient boosting (XGBoost)[32–33], light gradient boosting machine (LGBM)[35], Lasso and Ridge.
XGBoost. We use XGBoost model with 100 trees that randomly sample 90 percent of the training data prior to growing trees. We choose the dart booster as the XGBoost’s booster which can prevent overfitting and improve the model performance. We use the implementation provided the famous dmlc XGBoost’s Python Package with the above-mentioned parameters to train the model.

LGBM. We also use LGBM model to contrast with other model. The LightGBM model’s parameters are similar with XGBoost model, such as booster and subsample. However, we choose 2000 trees to construct our LGBM model with 0.01 learning rate. As for other parameters, we use the GridSearchCV algorithm which provided by scikit-learn to search the best parameters. We use the Python Package of LGBM to build our model.

Lasso and Ridge. Both the Lasso and the Ridge are liner models. They are usually used for feature selection. Lasso takes the L1 penalty for both fitting and penalization of the coefficients. Ridge takes the L2 penalty. They all have coefficients for every feature, which visually show correlation between the feature and the target. However, because of the difference of penalty, Lasso would be forces certain coefficients to zero and Ridge would only change the value without changing to zero. In our study, we also use the scikit-learn package. Besides, because of the processing of data with Min-max normalization, we do not normalize our data again and set the “cv” parameter to 5.

5. Evaluation

In this section, we analyze the effect of team competition on college students’ academic performance by answering the following research questions:

RQ1: How does different machine models perform in ITE prediction?

RQ2: Which features are most correlated with students’ academic performance when conducting team competition in college classroom?

RQ3: How does different competition design methods impact the effect of team competition on students’ academic performance?

5.1 Performance Comparison

Following the standard practice, we randomize the dataset and split it into training set, validation set and test set. We adopt RMSE, which is commonly used in measuring the accuracy of a machine learning predictor[35–37]:

\[
RMSE = \sqrt{\frac{\sum_{i=1}^{N}(ITE_i - \hat{ITE}_i)^2}{N}},
\]  

where \(N\) indicates the sample size.

The prediction accuracy of the models on both validation set and test set is illustrated in Figure 5. To test validity of our study, we construct two baselines. The random baseline retrieve a random value from a Gaussian distribution that is estimated by ITEs in the training set. The average baseline predicts all ITEs in the test set as the mean value of all ITEs in the training set. Figure 5 shows that XGBoost, LGBM, Ridge and Lasso all achieve similar accuracy, demonstrating significant advantage over average and random baselines in RMSE by up to 95% and 30%, respectively.
5.2 Analysis of Feature Importance

XGBoost, LGBM and Lasso can select features in training process. By removing features with zero coefficients in Lasso, and features with negative importance in XGBoost and LGBM, we can find out the salient predictors for the three models. Note that because of the difference in structure, different models may choose different features, as shown in Figure 6.

Figure 5. Comparison of model performance (RMSE)

Figure 6: Importance scores of features
We investigate the importance of features from different ITE prediction models. Figure 6 (a) and (b) show the selected feature from the Lasso and Ridge models. Figure 6 (c) and (d) illustrate the top 15 most important features selected from XGB and LGBM models.

The academic performance features of teams and individuals before the competition, e.g., average Aptis score of a team, the overall Aptis score and the Aptis score of four skills, overall score and scores of all subjects in NCEE, show strong predictive power in ITE prediction (see Figure 6). Surprisingly, average Aptis score of a team is the largest negative factor in both Lasso and Ridge model. The finding is consistent with the relationship between ITE and average team performance, as shown in Figure 7. Teams with the highest Aptis score yield smaller treatment effects than teams with low Aptis score. Moreover, the Aptis score of writing, speaking, listening and reading are also negative factors in Lasso and Ridge model. Moreover, they are import features in XGBoost. Their relationships with ITE is consistent with the relationship between the ITE and the average Aptis score of a team, which suggests that students with low academic performance may benefit more from the application of team competition in college English teaching.

![Figure 7: Relationship between average performance of a team and ITE](image)

5.3 Impact of Competition Design

The way of team formation is a significant predictor in ITE prediction. Figure 8 illustrates ITEs of three treatment groups that form teams in different methods and the ITE of the control group that does not conduct team competition. As we can see from Figure 8, self-formed treatment group obtains the biggest treatment effect. The result is consistent with the conclusion drawn in other domains [12]. The reason is that students from self-formed treat groups are usually acquaintances in real life, which may lead to higher level of team identity and responsibility. Grade-balanced treatment group yield smaller treatment effect than self-formed treatment group, but its treatment effect is bigger than the other two groups. The finding provides insights for team formation in scenarios when students are not familiar with each other. Not surprisingly, the treatment effect of control group is approximately to 0. A rather intriguing finding is that random-assigned treatment group obtains the smallest treatment effect, indeed, negative treatment effect.
In addition, we also investigate the average discussion times of each group, as shown in Figure 8 (b). We can find out that the number of discussions self-formed treatment group engaged in is the most, and the number of discussion random-assigned treatment group participate in is the least. The number of discussions that grade-balanced group participate is bigger than that of control group, but smaller than that of self-formed treatment group. This is consistent with the average ITE of the four groups. The result shows that self-formed group is more proactive than the other groups, and obtain the biggest individual treatment effect. Moreover, we can also conclude that introducing team competition into college English teaching may not necessarily have positive effect on students' academic performance, which depends on how team competition is conducted.

6 Conclusion

In conclusion, our research delved into two crucial realms: the impact of team competition on college students’ academic performance and the integration of Machine Learning techniques with MEC terminal data. Through rigorous randomized field experiments among college freshmen, we meticulously analyzed team-related and individual features, employing advanced machine learning models. Our findings underscored the significant predictive power of these features on academic performance, enabling a reduction in prediction errors by up to 30%. Moreover, our study provided valuable insights into the practical application of team competition strategies within college classrooms, offering immediate implications for the teaching design of college English. While our research represents a foundational step, further exploration is essential. Future endeavors will encompass additional field experiments, extending our insights to various courses, and addressing unresolved issues in the intersection of Machine Learning and MEC data processing. This interdisciplinary approach paves the way for enhancing educational methodologies, fostering active student engagement, and advancing the integration of cutting-edge technologies in contemporary learning environments.

Funding: This research was funded by Shandong Provincial Natural Science Foundation of China (GrantNo. ZR2021QG021) and Shandong Social Science Planning Fund Program of China (GrantNo. 22CSDJ19)

Conflicts of Interest: The authors declare no conflict of interest

Authorship definition: Conceptualization, H.Z. and Y.Y.; methodology, H.Z. and W.G.; software, Z.L.; validation, Z.L. and Y.Y.; formal analysis, H.Z.; investigation, H.W.; resources, H.Z.; data curation, H.W.; writing—original draft preparation, H.Z.; writing—review and editing, H.Z.; visualization, Z.L.; supervision, Y.Y. and W.G.; project administration, H.Z.; funding acquisition, H.Z. All authors have read and agreed to the published version of the manuscript.

References


