IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 13 October 2024, accepted 2 December 2024, date of publication 11 December 2024, date of current version 23 December 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3514694

RESEARCH ARTICLE

What Influences the Availability of STEM and Other Extracurricular Clubs in American High Schools?

AMY TANG[1], **YANG WANG**[2], **AND CHUNQIANG TANG**[3]
[1]Lynbrook High School, San Jose, CA 95129, USA
[2]Department of Computer Science, The Ohio State University, Columbus, OH 43210, USA
[3]Meta Platforms, Menlo Park, CA 94025, USA

Corresponding author: Amy Tang (amyzqtang@gmail.com)

ABSTRACT This paper provides a comprehensive categorization of extracurricular clubs (e.g., robotics, STEM, and business clubs) in hundreds of American high schools and examines how socioeconomic factors—such as race, income level, pupil-teacher ratio, and school size—affect the quantity and variety of these clubs. We reveal that robotics clubs have become the most prominent extracurricular activity in promoting electrical and computer engineering, as well as STEM fields at large, outperforming even math clubs, with 38.9% of schools hosting them. Although schools across different socioeconomic backgrounds all manage to offer robotics clubs, school demographics do affect the total number of clubs a school provides. Nevertheless, successful schools are capable of expanding their extracurricular offerings despite these constraints. Specifically, within schools of similar demographics, the top 25% offer 8.8 times more clubs than the bottom 25%.

INDEX TERMS American high school, extracurricular clubs, socioeconomic status, STEM, robotics.

I. INTRODUCTION

The National Survey of America's Families showed that 83% of youth aged 12 to 17 took part in at least one extracurricular activity within the preceding year [40]. With the increasing emphasis on STEM education and the encouragement for students to pursue careers in STEM fields, American high schools have extensively utilized STEM-related extracurricular clubs to support this initiative [53]. In particular, robotics clubs have emerged as the most popular and effective in promoting early education in electrical and computer engineering [26], [27], [37], [55].

The importance of high school extracurricular clubs in general, and STEM clubs in particular, is grounded in several key learning theories. The interest-based learning theory [12] emphasizes that students learn best when education aligns with their interests and involves active, hands-on experiences. The extracurricular and self-involvement nature of STEM clubs resonates well with participants' passions, and the success of hands-on activities further enhances their interest. The self-determination theory [11] posits that individuals are motivated to learn and grow when they experience autonomy, competence, and relatedness. STEM clubs empower students with the autonomy and agency to self-organize and pursue their goals, fostering a sense of ownership over their learning and propelling them toward success. The social learning theory [4] highlights the importance of observation and imitation in the learning process, while the situated learning theory [30] suggests that learning occurs through participation in communities of practice, where individuals engage in authentic activities. Aligned with these theories, STEM clubs create opportunities for students to observe, practice, and collaborate within a community environment, learning from peers and mentors. Finally, the expectancy-value theory [54] emphasizes that individuals' beliefs about their abilities and the value they assign to a task influence their motivation and performance. STEM clubs support this by offering opportunities for students to succeed—such as building a functioning robot—while also demonstrating the tangible values of STEM activities.

The associate editor coordinating the review of this manuscript and approving it for publication was Meriel Huggard.

In addition to the aforementioned theories, past research has provided concrete evidence of a strong association between positive youth development and active involvement in extracurricular activities [7], [10], [13], [14], [18], [19], [20], [21], [35], [36], [43], [45], [56]. Additionally, several surveys have summarized the effects of participation in extracurricular activities [15], [16], [24], [49]. A prominent taxonomy, known as "the five Cs," attributes the positive outcomes to the beneficial effect of organized extracurricular activities on five key areas of youth development: competence, confidence, connection, character, and caring [32].

Despite a large body of research on youth extracurricular activities, several notable gaps remain, which this study aims to address.

First, previous investigations have predominantly focused on the demand side of the student-school relationship, specifically student participation [17], [19], [28], [38]. In contrast, this study shifts the focus to the supply side—namely, whether certain clubs are offered by schools. For instance, if a school does not offer a robotics club, students interested in robotics may miss opportunities to further develop their passions. This missed opportunity will not be reflected in the student-participation data presented in prior studies.

Second, the absence of detailed data on club classifications and availability hinders effective policy-making. For instance, how are clubs distributed across categories such as career & skills, arts & creativity, STEM, hobbies & interests, and social issues? Do schools oversupply certain categories, like arts & creativity, while undersupplying others, like STEM? Is the LGBTQ+ rights movement sufficiently represented through relevant clubs? This study will address these kinds of questions, offering insights that can guide policy decisions.

Third, limited research exists on how the availability of specific club categories, such as robotics clubs, is influenced by students' socioeconomic factors, including race, income level, pupil-to-teacher ratio, and school size. For instance, it might be assumed that schools in lower-income areas are less likely to offer robotics clubs due to the associated equipment costs. Interestingly, this study will demonstrate that this assumption does not hold. More broadly, there is currently no data to confirm or refute similar hypotheses.

Finally, while past studies have tackled different questions, their methods generally relied on interviews or surveys, which are not scalable and can only collect data from a small number of schools. Even if these methods were applied to the questions in this study, the limited samples would not be statistically significant to represent American high schools. In contrast, this study leverages the key insight that, in the era of the internet, many schools already publish club data on their websites, though these data are often in different formats and require data cleansing. The shift in methodology to utilize online data enables this study to collect data from hundreds of schools to address previously unexamined questions.

In the following sections, we describe how this study addresses the above gaps identified in previous research, beginning with an overview of the methodology used.

II. METHODOLOGY AND DATA COLLECTION

We study American high schools, which typically cover grades 9 to 12. To focus on the most popular type of schools, we consider only public schools, deferring the examination of private schools to future research.

This study requires school demographics data, which we obtain from the American National Center for Education Statistics (NCES) [42]. The data we utilize are school sizes, free or reduced-price lunch statistics as indicators of income levels, pupil-to-teacher ratios, and racial demographics. Previous studies have often considered these factors in socioeconomic research, and we continue this tradition.

Considering that a school's location—such as its city, school district, longitude, and latitude—may influence its social profile, we initially included it as a factor in our study. However, we discovered that incorporating location did not provide more meaningful insights. For instance, two schools located in the same city and school district, with similar longitude and latitude, may display significantly different social profiles due to differences in income levels. Instead, school profiles are often better represented by direct factors we already use, such as free or reduced-price lunch statistics, pupil-to-teacher ratios, and racial demographics.

Additionally, we exclude gender as a factor in the study because most schools have balanced enrollment by gender; consequently, it has no strong correlation with the results. We also exclude grades as a factor in this study because the club data is often not categorized by grade; sometimes, students from different grades can participate in the same club.

In addition to school demographics, this study also requires data from a large number of American high schools regarding the complete list of extracurricular clubs offered by each school. Traditionally, researchers would obtain such data through direct interviews or surveys at specific schools. However, this approach is not only labor-intensive but also necessitates individual approval from each school involved, making it challenging to scale to hundreds of schools across tens of states in the United States. Instead, we note that many schools already publish club data on their websites. By employing careful data selection and cleansing techniques, we can transform this information into high-quality data suitable for research purposes.

Concretely, starting with the schools listed in the NCES dataset, we randomly sampled thousands of American public high schools and explored their websites. Although the vast majority of these schools provide some club information on their websites, many display only a few examples of clubs rather than the entire list. Overall, less than 10% of schools provide sufficiently complete club information. After manually exploring the websites of thousands of schools, we filtered out most and identified 229 remaining schools whose data are of sufficiently high quality, which we use in this study. Examples of these websites can be found in [8] and [9].

This study excludes clubs for varsity sports such as swimming and baseball for multiple reasons. First, they warrant a separate, dedicated study due to their substantial presence. Second, school websites treat varsity sports inconsistently, some as clubs and others as athletics. Finally, information about common varsity sports like basketball is often missing from school websites, despite the high likelihood of them being offered by schools. To ensure consistency, we exclude varsity sports from this study, leaving them for future research. However, intramural sports like mountain biking are included in this study as they are consistently listed as clubs.

After compiling the complete list of a school's clubs, we manually classify each club into categories such as STEM or arts & creativity. We then conduct statistical analyses to explore the relationships between schools' demographic data (e.g., school size and racial composition) and club data (e.g., the number, type, and variety of clubs offered). All data processing and statistical analyses are carried out using a Python program we developed, which utilizes the open-source Python statistical package `statsmodels` [50].

III. ANALYZING CLUB DATA

In this section, we first manually categorize clubs and then construct models to predict the number of clubs in specific categories, such as STEM.

A. CLUB CATEGORIZATION

We manually classified a total of 5,983 clubs offered by 229 schools. 2.2% of those clubs cannot be categorized due to ambiguous club names and insufficient descriptions, such as clubs named "Sharks United." The remaining clubs are mapped to 1,250 unique clubs, by combining identical ones (e.g., local chapters of Key Club) or closely related clubs (e.g., slight variations of tabletop games) into a single entity.

We further map these 1,250 unique clubs into the 17 categories shown in Table 1. Note that these categories are not mutually exclusive. For instance, a robotics club is listed under multiple categories in the table, including *robotics*, *STEM*, *career & skills*, and *computer science*.

A brief explanation of the caveats in defining the club categories follows. The *career & skills* category includes clubs directly tied to career preparation for specific professions, as well as those emphasizing the development of crucial and general skills, like speech and debate. The *hobby & interest* category encompasses specific hobbies as well as activities rooted in broad interests that are challenging to place elsewhere, such as philosophy and travel. The *culture, identity & religion* category broadly addresses cultural and group identity matters. The rows in the table below the *intramural sports* row denote specific categories rather than general ones. We still include them because they offer interesting perspectives for study. The *intramural sports* category reflects a caveat we discussed earlier—this study excludes varsity sports but includes intramural sports.

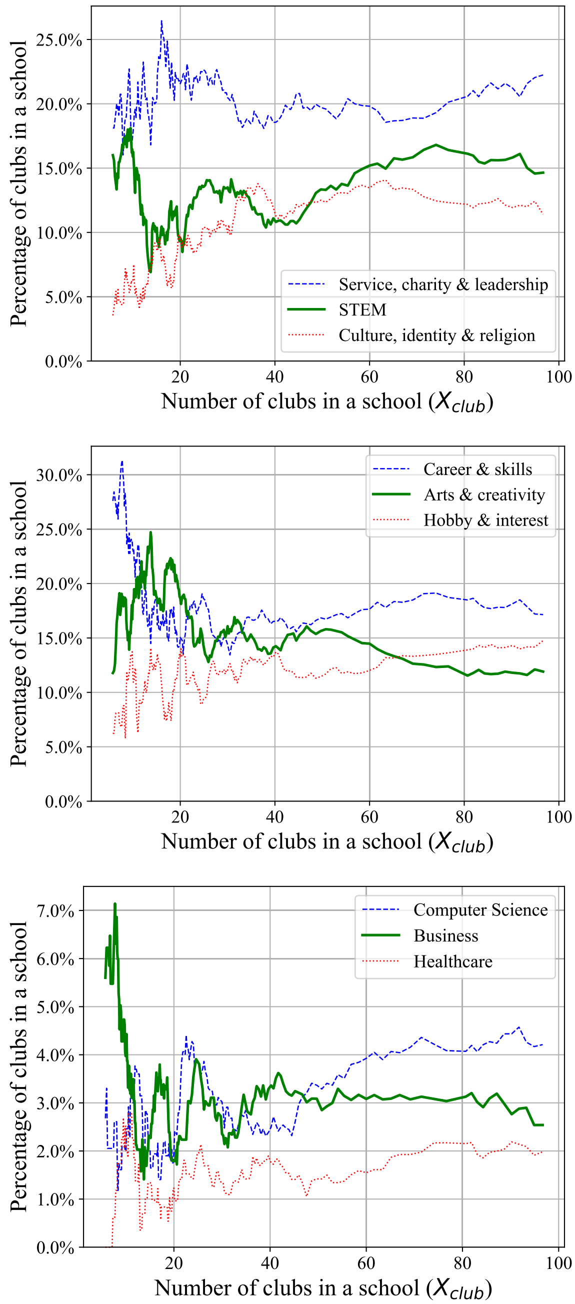


FIGURE 1. These figures show how the percentages of clubs belonging to specific club categories change as the number of clubs (X_{club}) increases. Each curve is computed as a moving average across schools.

Table 1 reports the club categories by club count, school count, and student count. Let's illustrate how to read it using an example. The second row of the table indicates that the *service, charity, and leadership* category represents *20.6%* of the total 5,983 clubs across the 229 schools. Clubs in this category are offered by *92.6%* of the 229 schools, and these schools collectively accommodate *98.2%* of the 221,300 students enrolled in the 229 schools.

TABLE 1. Coverage of club categories by club count, school count, and student count, respectively.

Club category	Examples	By club count	By school count	By student count
Service, charity & leadership	Key Club, Interact Club, environmental club, charity for cancer, school newspaper, math tutoring, community service	20.6%	92.6%	98.2%
Career & skills	Future Farmers of America (FFA), Health Occupations Students of America (HOSA), computer science, business, debate	17.4%	95.6%	98.5%
Arts & creativity	Art club, drama, music band, chorus, thespian, fantasy writing	15.5%	91.7%	96.8%
STEM	Robotics, math club, FFA, HOSA, Science Olympiad	13.1%	89.1%	95.4%
Hobby & interest	Chess, book club, Esports, tabletop game, anime	12.0%	75.5%	90.3%
Culture, identity & religion	Black Student Union, multiculture club, traditional Chinese clothing, Fellowship of Christian Athletes	10.5%	74.7%	87.5%
Academic	Math club, National Honor Society (NHS), Academic Decathlon, chemistry	10.0%	90.4%	93.4%
Social issues	LGBTQ+, destructive decision awareness, women empowerment, Junior State of America, Young Democrats of America, pro-choice/pro-life	7.8%	63.8%	81.1%
Honor society	National Art Honor Society, National Spanish Honor Society, NHS	6.2%	74.7%	82.1%
Intramural sports	Frisbee, trap shooting, mountain biking	3.3%	46.3%	54.3%
Computer Science	Robotics, Girls Who Code, Cyberpatriot, game design, artificial intelligence	3.2%	48.5%	65.5%
Business	Future Business Leaders of America (FBLA), Distributive Education Clubs of America (DECA), investment, finance, entrepreneurship	3.0%	54.1%	65.1%
LGBTQ+	Gay Straight Alliance, LGBTQ+, gender equality	1.9%	47.2%	66.1%
Math	Math club, Math Honor Society, calculus, statistics, math tutoring	1.7%	35.4%	47.6%
Robotics	Robotics	1.6%	38.9%	52.3%
Healthcare	HOSA, pre-med, neuroscience, sports medicine, future veterinarians	1.5%	30.1%	48.8%
Agriculture	FFA, greenhouse and vocational agriculture	1.5%	37.6%	35.7%

TABLE 2. Summary of variables used in this study.

Variable	Explanation
X_{enroll}	School size, i.e., student enrollment.
X_{teacher}	Pupil/teacher ratio. A higher X_{teacher} means fewer teachers per student.
X_{lunch}	Fraction of students in a school receiving free or reduced-price lunch. A higher X_{lunch} value indicates a lower household income.
X_{race} (e.g., X_{white})	Fraction of students of a specific race in a school.
X_{club}	The total number of clubs offered by a school.
X_{stem}	The number of STEM clubs offered by a school.

We draw several observations from the table. First, while certain categories are more popular than others, the distribution of club categories is diverse, with no strong dominance by one or two categories.

Second, when considering the coverage by school count, the variety of club categories becomes even more apparent compared to just examining the club count. For instance, although the *STEM* category's representation by club count is notably lower than that of the *service* category, they exhibit nearly identical representation by school count—89.1% versus 92.6%. This indicates that despite having fewer STEM clubs, a comparable percentage of schools offer them, ensuring students interested in STEM have access. Moreover, in terms of student count, STEM clubs encompass schools enrolling 95.4% of all students.

Third, the most substantial subcategories within the *career* category—namely *computer science*, *business*, *healthcare*, and *agriculture*—contribute small proportions by club count (ranging from 1.5% to 3.2%). However, they exhibit significant coverage by school count (ranging from 37.6% to 48.5%), and even more so by student count (ranging from 35.7% to 65.5%).

Fourth, although the most popular major in colleges is business [3], STEM outperforms it in terms of high school club offerings, showing positive emphasis on STEM in high schools.

Fifth, surprisingly, the *robotics* club stands out as the most popular STEM club, even more popular than math clubs, covering 38.9% of schools and 52.3% of students. Later, we will specifically examine the factors that impact its availability.

Sixth, the *math* category reaches only 35.4% of schools and 47.6% of students, which is low considering the pervasive role of math in education. Compared with other STEM subjects, the popularity of math clubs is concerning.

Lastly, the *LGBTQ+* category exhibits extensive coverage by school count (47.2%) and even higher coverage by student count (66.1%). This demonstrates that the LGBTQ+ rights movement has rapidly advanced in reaching schools and students.

B. SHIFT OF CLUB COMPOSITION

After analyzing the breakdown of club categories across all schools, one remaining question is whether the composition of club categories varies for schools with different numbers of clubs. Furthermore, when a school effectively increases its club count through targeted endeavors, would the mix of its club categories also undergo a change?

To simplify the description of our answers to these questions, we utilize the variables defined in Table 2. Specifically, we graph the percentages of clubs in certain categories against the club count (X_{club}) in Figure 1. Let's use the *STEM* curve in the first figure as an example to elucidate how it is constructed. First, for each school, we calculate the percentage of its STEM clubs out of its total clubs, denoted as $P_{\text{stem}} = X_{\text{stem}}/X_{\text{club}}$. Then, we sort the schools by their number of clubs (X_{club}). For each group of 10 consecutive schools within the sorted list, which have similar X_{club} values, we compute the average of their X_{club} values denoted as $\bar{X}_{\text{club}}$, as well as the average of their P_{stem} values, denoted as $\bar{P}_{\text{stem}}$. The data point $(\bar{X}_{\text{club}}, \bar{P}_{\text{stem}})$ represents the average values of the 10 schools, and is plotted as one data point on the *STEM* curve. By repeating this procedure for all groups of 10 consecutive schools within the list sorted by X_{club}, we can plot the entire moving-average curve for *STEM*. The moving-average curves for other categories are constructed in a similar fashion.

Figure 1 demonstrates that as X_{club} increases, the proportions of clubs in different categories show no drastic change. One noteworthy exception is the *culture* category. Initially, with increasing X_{club}, the fraction of *culture* clubs exhibits steady growth, leveling off after X_{club} surpasses 40. Another encouraging trend is the slight increase in the proportions of *STEM* and *career* clubs as X_{club} exceeds 40, suggesting ample attention to career preparation.

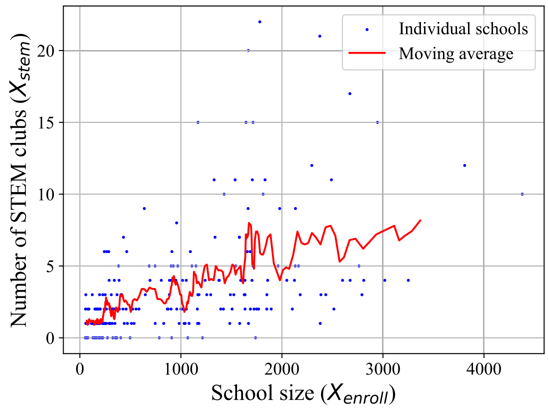
C. MODELING STEM CLUBS

After presenting a breakdown of club categories, we analyze how different factors influence the availability of clubs in specific categories. Our goal is to develop a model for predicting the number of clubs in specific categories within a school. Given society's focus on promoting STEM education, we initiate our analysis with the STEM category. Refer to Table 2 for the definitions of the variables.

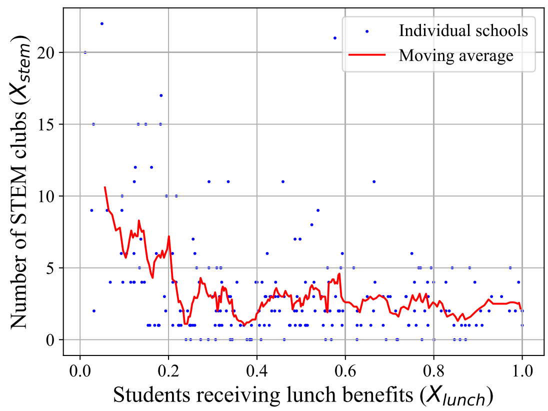
To gain an initial intuitive understanding, we first visualize how different factors impact the count of STEM clubs (X_{stem}) offered by schools. Figure 2a depicts the relationship between X_{stem} and school size (X_{enroll}) for each of the 229 sampled schools. Each dot represents a school, while the "*moving average*" curve portrays the overall trend. The moving-average curve is constructed similar to the curves in Figure 1.

Similarly, in Figure 2b, we plot the relationship between X_{stem} and the fraction of students receiving free or reduced-price lunch (X_{lunch}). This figure seems to indicate a negative correlation between X_{stem} and X_{lunch}. When X_{lunch} is less than 40%, as it increases, implying a decrease in household income, X_{stem} decreases accordingly. This suggests that lower-income schools tend to offer fewer STEM clubs. As X_{lunch} further increases above 40%, its impact on X_{stem} flattens out.

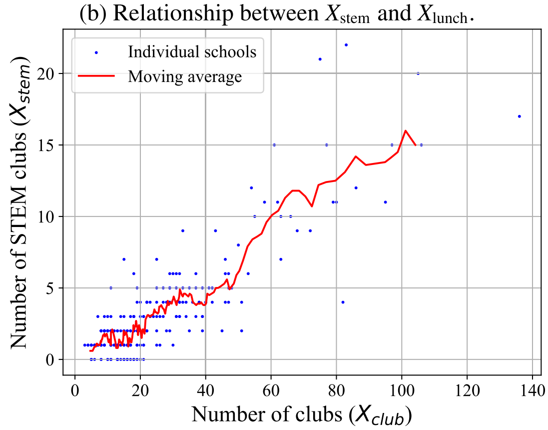
Finally, Figure 2c plots the relationship between X_{stem} and the total number of clubs (X_{club}) offered by a school. This figure suggests that as schools offer more clubs, they also tend to proportionally offer more STEM clubs. Specifically, X_{club} and X_{stem} exhibit a correlation coefficient of 0.86 and $p < 0.001$.



(a) Relationship between X_{stem} and X_{enroll}.



(b) Relationship between X_{stem} and X_{lunch}.



(c) Relationship between X_{stem} and X_{club}.

FIGURE 2. These figures show the relationship between the number of STEM clubs (X_{stem}) and X_{enroll}, X_{lunch}, and X_{teacher}, respectively. Each dot represents one school, and the moving average curves show the overall trend.

Based on the above observations, we formulate the model for X_{stem} as follows:

$$X_{\text{stem}} = \beta_0 + \beta_{\text{club}} X_{\text{club}} + f(X_{\text{enroll}})[\beta_{\text{enroll}} + \beta_{\text{lunch}} g(X_{\text{lunch}}) + \beta_{\text{teacher}} X_{\text{teacher}} + \sum_{\text{race}} \beta_{\text{race}} X_{\text{race}}] + \epsilon \tag{1}$$

where

$$f(X_{\text{enroll}}) = \sqrt{X_{\text{enroll}}}$$
$$g(X_{\text{lunch}}) = \frac{1}{1 + e^{-K(X_{\text{lunch}} - X_0)}} \tag{2}$$

Applying the Levenberg-Marquardt algorithm to fit the data, we obtain the parameters $K = 13.8$ and $X_0 = 0.049$ for $g(X_{\text{lunch}})$.

The function $f(X_{\text{enroll}})$ adopts a square-root form due to the relatively slower growth of X_{club} compared to X_{enroll}. Meanwhile, $g(X_{\text{lunch}})$ is modeled as a logistic function to capture the observed flattening-out effect in the data. Specifically, once X_{lunch} surpasses the threshold of 40%, its impact is capped, effectively behaving as if its value were 40%, as shown in Figure 2b.

A unique aspect of this model is the use of $f(X_{\text{enroll}})$ as a multiplying factor for X_{lunch}, X_{teacher}, and X_{race}. This choice is due to the influential role of school size. As $f(X_{\text{enroll}})$ approaches zero, indicating a diminishing student population, X_{stem} should tend towards zero, regardless of X_{lunch}, X_{teacher}, and X_{race}. Additionally, school size amplifies the impact of other independent variables. For example, assuming higher income levels contribute to more clubs, this effect should yield a more significant absolute increase in club counts for larger schools.

We evaluate the prediction accuracy of the model within two distinct scenarios: one with the X_{club} term, and the other without it. In the absence of the X_{club} term, the model represents a scenario in which only basic school demographics are available, while the precise club count (X_{club}) remains unknown. This scenario has wider applicability as it requires fewer inputs. However, since a school's demographics (X_{enroll}, X_{lunch}, X_{teacher}, and X_{race}) undergo minimal changes over short time spans, the insights derived from the model without X_{club} may offer limited immediate utility to school administrators.

Conversely, the club count (X_{club}) can potentially be increased swiftly through focused initiatives, as demonstrated by successful examples from the past [45]. Hence, our interest lies in evaluating the impact of changes in X_{club} —while keeping X_{enroll}, X_{lunch}, X_{teacher}, and X_{race} constant— on specific club categories such as X_{stem}. These insights might hold more immediate practical value for school administrators.

1) PREDICTING X_{STEM} WITHOUT X_{CLUB}

The regression results for the scenario without the X_{club} term is summarized in Table 3. We use the linear regression implementation in the `statsmodels` package [50] to compute the coefficients. Due to the presence of heteroscedasticity in the residual, we employ the HC3 covariance matrix estimator in `statsmodels` to calculate heteroskedasticity-robust standard errors. This estimator implements the algorithm proposed by MacKinnon and White [33].

TABLE 3. This table shows the regression results for the model in Equation 3, which predicts X_{stem} based on X_{enroll}, X_{lunch}, and X_{teacher}. The model has an adjusted R^2 of 0.4540, with an F-statistic of 27.03 and $p < 10^{-14}$. β_{teacher} is noticeably smaller than β_{lunch} for two reasons. First, the scale of X_{teacher} (mean: 15.3) is much larger than that of X_{lunch} (mean: 0.5). Second, X_{teacher} has a smaller impact on X_{stem}.

Coefficient symbol	Coefficient value	Robust standard error	t-value	p-value	Confidence interval [0.025	0.975]	Partial R^2
β_0	-0.7806	0.53	-1.472	0.142	-1.825	0.264	0.0118
β_{enroll}	0.8222	0.155	5.312	< 0.001	0.517	1.127	0.3276
β_{lunch}	-0.3057	0.087	-3.528	0.001	-0.476	-0.135	0.2079
β_{teacher}	-0.005	0.002	-2.382	0.018	-0.009	-0.001	0.0509

The signs of the coefficients in Table 3 indicate that larger schools, higher-income schools, and schools with lower pupil/ratios tend to offer more STEM clubs. The magnitude of the partial R^2 values indicates that their impact on X_{stem} is ranked in the order of X_{enroll}, X_{lunch}, and X_{teacher}.

When the model employs solely X_{enroll}, X_{lunch}, and X_{teacher} (excluding X_{club} and X_{race} terms) to predict X_{stem}, it yields an adjusted R^2 of 0.4540. Adding the X_{white} variable to the model, without any other X_{race} variables, only negligibly improves the adjusted R^2 to 0.4571. Finally, adding all X_{race} variables into the model leads to multicollinearity due to their inherent summation to one and strong correlations. Hence, removing at least one X_{race} variable is necessary. However, eliminating an X_{race} variable associated with a racial group of small school population inadequately addresses multicollinearity, as other high-value X_{race} variables maintain strong correlations. One effective solution involves the removal of the X_{race} variable with the highest value, i.e., X_{white}. With X_{white} eliminated and multicollinearity resolved, the inclusion of additional X_{race} variables, regardless of their combination, never elevates the adjusted R^2 beyond 0.4585, indicating negligible enhancement over the simpler model.

The above observations indicate that the inclusion of X_{race} variables complicates the model without improving its data fit. Therefore, for the scenario without the X_{club} term, we use the following simplified model that excludes the X_{race} terms:

$$X_{\text{stem}} = \beta_0 + f(X_{\text{enroll}})[\beta_{\text{enroll}} + \beta_{\text{lunch}} g(X_{\text{lunch}}) + \beta_{\text{teacher}} X_{\text{teacher}}] + \epsilon \tag{3}$$

2) PREDICTING X_{STEM} WITH X_{CLUB}

Next, we examine the scenario of adding the X_{club} term to the model. The results are summarized in Table 4. Despite X_{stem}'s apparent correlation with X_{enroll} and X_{lunch} in Figures 2a and 2b, further investigation reveals that when X_{club} is known, it becomes the sole determining factor influencing X_{stem}. In other words, if two schools offer the same number of clubs (X_{club}), regardless of their different values in X_{enroll}, X_{lunch}, and X_{teacher}, they tend to offer the same number of STEM clubs (X_{stem}) as well. When X_{club} is unknown, the impact of X_{enroll}, X_{lunch}, and X_{teacher} on X_{stem} appears due to their influence on X_{club}, subsequently exerting an indirect effect on X_{stem} through X_{club}.

Concretely, using X_{club} as the sole predictor without other variables yields an adjusted R^2 of 0.7323. Further adding X_{enroll}, X_{lunch}, and X_{teacher} to the model regresses the adjusted R^2 to 0.7309. Moreover, the partial R^2 values for X_{club}, X_{enroll}, X_{lunch}, and X_{teacher} are 0.5093, 0.0045, 0.0050, and 0.0011, respectively, reflecting that the impact of X_{enroll}, X_{lunch}, and X_{teacher} is negligible.

Hence, for the scenario with the X_{club} term, we use the following simple model that excludes all other variables:

$$X_{\text{stem}} = \beta_0 + \beta_{\text{club}} X_{\text{club}} \tag{4}$$

The regression results for this simple model is summarized in Table 4.

TABLE 4. Regression results for the model in Equation 4, which predicts X_{stem} based on X_{club} alone. The model has an adjusted R^2 of 0.7323, with an F-statistic of 177.6 and $p < 1^{-29}$.

Coefficient symbol	Coefficient value	Robust standard error	t-value	p-value	Confidence interval		Partial R^2
					[0.025	0.975]	
β_0	0.1392	0.051	2.746	0.007	0.039	0.239	0.0356
β_{club}	0.0136	0.002	7.537	< 0.001	0.010	0.017	0.2849

D. SCHOOL INITIATIVES IN IMPROVING STEM CLUB OFFERINGS

While school demographics, such as school size, household income, and pupil/teacher ratio, do impact the number of STEM clubs (X_{stem}), we want to understand to what extent school initiatives within these demographic constraints can alter the outcome. For this purpose, we compare the top and bottom-performing schools regarding STEM club offerings.

For each school, we calculate the ratio, denoted $R_{\text{initiative}}$, between its actual STEM club count and the predicted count based on X_{enroll}, X_{lunch}, and X_{teacher} (Equation 3). A higher $R_{\text{initiative}}$ indicates superior performance given the constraints of a school's demographics. We rank schools according to $R_{\text{initiative}}$, and compare the top group for the best-performing 25% of schools and the bottom group for the worst-performing 25% of schools. A t-test between these two groups, with $p < 0.001$, shows a significant difference in their $R_{\text{initiative}}$ value. Specifically, the median $R_{\text{initiative}}$ for the top 25% of schools is 185%, while that for the bottom 25% of schools is only 21%. Intuitively, it means that among schools with identical demographics, the top schools have 8.8 times more clubs than the bottom schools. This substantial difference underscores the pivotal role of school initiatives in enhancing STEM club offerings.

Similarly, for each school, we calculate the ratio ($R'_{\text{initiative}}$) between its actual STEM club count and the predicted count based on X_{club} alone (Equation 4), and compare the top-performing 25% of schools with the bottom-performing 25% of schools in terms of their $R'_{\text{initiative}}$ value. A t-test between these two groups, with $p < 0.001$, shows a significant difference in their $R'_{\text{initiative}}$ value. Specifically, the median $R'_{\text{initiative}}$ for the top 25% of schools is 201%, while that for the bottom 25% of schools is only 30%. This 6.7-fold difference, again, emphasizes the importance of school initiatives in enhancing STEM club offerings.

E. MODELING ALL CLUB CATEGORIES

Having constructed the models presented in Equations 3 and 4 to assess the impact of various factors on STEM clubs, we now apply these models to other club categories, by changing the dependent variable, X_{stem}, in those equations to variables representing other club categories such as X_{service} and X_{career}. The regression results for all club categories are summarized in Table 5.

Upon comparing Tables 1 and 5, certain patterns and unexpected findings emerge. First, across the board, there is a higher prediction accuracy for popular categories that constitute a significant portion of clubs, including *service, charity & leadership*, *career & skills*, *arts & creativity*, *STEM*, *hobby & interest*, and *culture, identity & religion*. This phenomenon is expected because popular clubs have more data points for regression. Second, the *academic* category stands as an exception. Despite being fairly popular and accounting for 10% of clubs, its prediction accuracy is lower than even less popular categories such as *social issues* and *computer science*. Lastly, in general the prediction based on X_{club} is more accurate, but X_{club} still cannot accurately predict some categories such as *math* clubs. This implies that while efforts to increase the total club count (X_{club}) may accordingly boost club counts in many club categories, it may not directly help math clubs due to the lack of a strong correlation between X_{club} and math clubs. Therefore, dedicated efforts are needed to boost math clubs.

TABLE 5. This table shows the adjusted R^2 of models that predict the number of clubs in specific categories based on X_{club} alone (using Equation 4), or based on X_{enroll}, X_{lunch}, and X_{teacher} (using Equation 3).

	Adjusted R^2 of prediction based on X_{club} (using Equation 4)	Adjusted R^2 of prediction based on X_{enroll}, X_{lunch}, and X_{teacher} (using Equation 3)
Service, charity & leadership	0.7846	0.4824
Career & skills	0.7464	0.5209
STEM	0.7323	0.4540
Social issues	0.7109	0.4122
Hobby & interest	0.7062	0.4318
Culture, identity & religion	0.6912	0.4602
Arts & creativity	0.6281	0.3304
Computer Science	0.5629	0.3186
Academic	0.4754	0.3364
Business	0.403	0.3156
Healthcare	0.2951	0.2512
LGBTQ+	0.2817	0.2430
Honor society	0.2438	0.2903
Intramural Sports	0.2432	0.1669
Math	0.2283	0.1672
Robotics	0.1662	0.1173
Agriculture	0.0076	0.0124

F. FACTORS INFLUENCING ROBOTICS CLUBS

Intuitively, one might expect that larger, wealthier schools or those with more teachers are more likely to offer robotics clubs. However, the data reveals no statistically significant correlation between the chance of a school offering a robotics club and its values in X_{enroll}, X_{lunch}, and X_{teacher}. Furthermore, the "Robotics" row in Table5 demonstrates that, based on the values of X_{club}, X_{enroll}, X_{lunch}, and X_{teacher}, Equations 3 and 4 are not accurate predictors of whether a school offers a robotics club. This is intuitively visualized in Figure 3, which shows that schools with diverse values in X_{enroll} and X_{lunch} all offer robotics clubs. A contributing factor to the widespread popularity of robotics clubs is the increased availability of low-cost robotics education kits, with many priced under \$100, making them accessible to most schools.Overall, this is an encouraging trend, suggesting that schools, despite demographic limitations, can and do successfully offer robotics clubs, thereby providing valuable electrical and computer engineering education opportunities to more students.

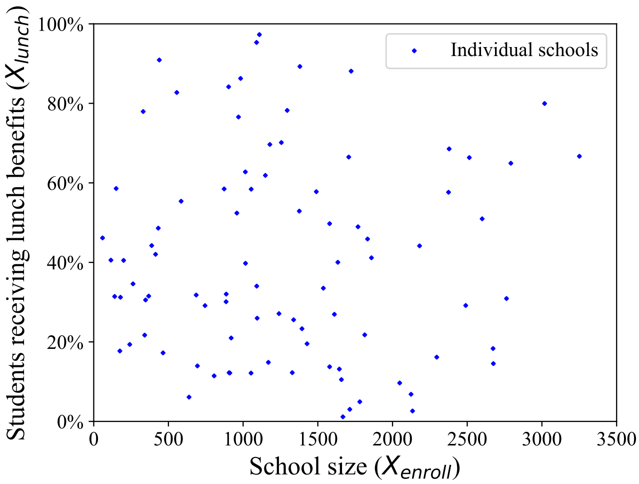


FIGURE 3. Schools offering a robotics club. Each dot in the figure represents a school, with its X_{enroll} and X_{lunch} on the x and y axes, respectively.

IV. DISCUSSION AND RECOMMENDATION

In this section, we present recommendations for policymakers and school administrators based on findings from our data analysis. First, robotics clubs have emerged as the most successful extracurricular activity in promoting electrical and computer engineering, as well as STEM fields in general. Although 38.9% of schools currently offer robotics clubs, there is still considerable room for expansion and improvement.

Our second recommendation is that schools should aggressively prioritize increasing the number of club offerings to boost student participation in positive extracurricular activities. While other concerns like club category composition, equitable STEM club coverage, or avoidance of concentration in narrow areas such as hobbies may arise, they should not take priority. Data consistently indicates that as clubs multiply, they naturally diversify across various areas, achieving a high club variety score, and organically covering crucial areas such as STEM, aligned with policymakers' and administrators' interests.

The statement from Principal Corey Tafoya of Woodstock High School, which had successfully improved the student participation rate in extracurricular activities by over 400% in five years, encapsulates our above recommendation most effectively: "*If we have six or seven students interested in something, we'll start a new club. We want students to find a reason to get up and come to school. Whatever trips their trigger is what our teachers and administration are willing to do*" [45]. Another successful example is Syracuse Academy of Science Charter School [52], which, despite its small enrollment of 286 students, with 75% receiving free or reduced-price lunch, offers an impressive 29 clubs—several times more than comparable schools.

Our third recommendation pertains to boosting extracurricular activities in math. Despite the pivotal role of math in education, math clubs are only present in 35.4% of the sampled schools, a rate even lower than that of robotics clubs. Departing from the common practice of math clubs concentrating on math competitions, we advocate a transition toward practical applications of math. Successful examples of clubs we have observed include applying math to sports/esports analytics, stock markets, and financial decisions.

V. LIMITATIONS OF THIS STUDY

This study has several limitations. First, this study exclusively focuses on schools that provide detailed club information on their websites, which may introduce bias, as better-equipped schools might be more likely to have comprehensive data available online. To verify this, one could conduct direct interviews or surveys with a smaller sample of schools and cross-validate the results with this study.

Second, the club data on the websites may be outdated and may not reflect reality. Conducting audits by interviewing a subset of the 229 investigated schools would provide insights into the accuracy of the data.

Third, if two schools offer the same club (e.g., a robotics club), this study treats them as achieving the same effect, even though one school's robotics club may have significantly more members than the other. This variation in membership is not considered in this study. We hypothesize that a club's membership size correlates with the square root of the school size, but this requires further validation.

Fourth, this study excludes private schools and varsity sports clubs, which may affect some of the conclusions. These present opportunities for future research.

Finally, while we have exercised our best judgment in the manual classification of club categories presented in Table 1, the various classification methods have not been thoroughly compared. This could also be a topic for future research.

VI. FUTURE WORK

In addition to addressing the limitations described in the previous section, we identify several opportunities for future research.

First, our analysis indicates that among schools with similar demographics, such as school size and income level, the top 25% offer 8.8 times more clubs than the bottom 25%. Understanding the policies or environmental factors that contribute to these significant differences is crucial.

Second, our data indicates that while STEM clubs are significantly more popular than business clubs in high schools, business is the most common college major, reversing their relative rankings. Further research is needed to investigate the relationship among college major selection, high school club availability and participation, and school demographics.

Third, our comparison of schools with varying numbers of club offerings suggests that as a school increases its clubs, the new additions tend to diversify across different categories, including STEM. However, to validate this finding, a longitudinal study is needed to examine the evolution of clubs within a specific set of schools.

Finally, the club names and descriptions we collected from the websites are in natural language. With the rise of AI technologies, natural language processing (NLP) can automate much of the analysis and provide deeper insights. Rather than manually classifying the club categories presented in Table 1, we can apply text clustering algorithms to objectively define club categories and automatically assign clubs to them. Additionally, by utilizing word embeddings from GloVe [44] or large language models like LLaMA [1], we can quantitatively assess the similarity between a pair of clubs, like robotics/astronomy or robotics/FBLA. By assessing the similarity among all of a school's clubs, we can define a numeric score that quantifies the variety of the school's club offerings, further examining our conclusion that American high schools provide a diverse range of club options.

VII. RELATED WORK

Prior research has presented compelling evidence regarding the robust connection between positive youth development and active engagement in extracurricular activities. This is substantiated by a range of studies [6], [7], [10], [13], [14], [17], [18], [19], [20], [21], [32], [35], [36], [38], [43], [45], [56]. Furthermore, multiple surveys have synthesized the effects of participating in extracurricular activities [15], [16], [24], [49].

Some research suggests that the benefits of extracurricular activities depend on the type of activities in which youth participate [29], [36]. Fredricks and Eccles [19] examined the impact of the total number and breadth of participation in activities on youth development. These studies rely on the assumption that a reasonable number and variety of extracurricular clubs are available to students. This research examines this assumption and studies the factors influencing the variety of high school clubs.

Due to society's strong focus on promoting STEM education, it has been widely studied in different contexts, including inclusive STEM high schools [39], class-based STEM learning [23], and out-of-school STEM education programs [5]. Unlike these works, this research studies the availability of school-sponsored STEM clubs. Jackson et al. introduced an equity-based framework of STEM literacy [25]. We find that an effective strategy for enhancing the accessibility of STEM clubs for students involves emphasizing broad policies that increase the overall number of clubs. These policies need not be exclusively oriented toward STEM. Within the broad STEM category, robotics has been long leveraged for STEM education [26], [27], [37]. Our analysis finds that this approach has been very effective in driving the learnings of subjects related to STEM in extracurricular clubs. In contrast, schools are yet to find such an effective driver for extracurricular math activities.

Barker and Gump's study [6] on school size and available extracurricular activities is relevant to this study. However, their work was limited to data from a small number of schools in a specific region (13 high schools in Eastern Kansas) and did not give a comprehensive classification of clubs.

Concerns about the over-scheduled child problem have been raised in some studies [46]. However, Mahoney et al. conducted an extensive survey and supported promoting participation in extracurricular activities, as they found limited empirical support for the over-scheduling hypothesis and consistent evidence for positive youth development [34].

Various studies have discussed the effects of large school sizes [31], [51], including their impact on student indiscipline [22], dropout rates [2], voluntary participation [48], social participation [41], and social networks [47]. While recognizing the significant long-term impact of factors like school size, our recommendation prioritizes short- to medium-term actions that can effectively increase club counts because both data and anecdotes suggest that there are ample opportunities for substantial improvements within the constraints of these long-term factors.

VIII. CONCLUSION

Using data collected from hundreds of American high schools, we have analyzed various statistics on extracurricular clubs and how school demographics influence the quantity and variety of clubs. We draw the following conclusions:

First, schools should focus on general policies that increase the total number of club offerings to boost student participation, rather than targeting specific categories or intentionally balancing club types. Data consistently show that as the number of clubs grows, they naturally diversify, covering key areas like STEM.

Second, although the most popular college major is business [3], STEM significantly outperforms it in terms of high school club offerings, showing a positive emphasis on STEM in high schools.

Third, robotics clubs, surpassing even math clubs, are the most popular STEM clubs, present in 38.9% of schools and serving 52.3% of the student population. Their availability shows no strong correlation with school demographics like school size or income level, likely due to the accessibility of low-cost robotics kits. This suggests that schools have found an effective way to promote STEM education, particularly in electrical and computer engineering.

Fourth, clubs in American high schools exhibit a wide variety, which is a positive sign. The most popular club categories include service, charity & leadership (20.6%), career & skills (17.4%), arts & creativity (15.5%), STEM (13.1%), hobby & interest (12.0%), culture, identity & religion (10.5%), academic (10.0%), and social issues (7.8%). Specifically, LGBTQ+ clubs exhibit extensive coverage by school count (47.2%) and even higher coverage by student count (66.1%). This demonstrates that the LGBTQ+ rights movement has rapidly advanced in reaching schools and students.

Fifth, three demographic factors—school size, income level, and pupil/teacher ratio—affect a school's club offerings, ranked in that order of significance. For schools with similar values in these factors, differences in racial composition show no statistically significant impact on club offerings. However, racial composition may still have an indirect effect, as it correlates with the three major factors.

Finally, while school demographics do influence the number of clubs a school offers, schools can and do enhance their club offerings despite these limitations. Specifically, among schools with similar demographics, the top 25% offer 8.8 times more clubs than the bottom 25%.

REFERENCES

[1] A. Grattafiori et al., "The llama 3 herd of models," 2024, *arXiv:2407.21783*.
[2] J. W. Alspaugh, "The relationship of school-to-school transitions and school size to high school dropout rates," *High School J.*, vol. 81, no. 3, p. 154, 1998.
[3] (2021). *Bachelor Degree Breakdown*. [Online]. Available: https://nces.ed.gov/programs/digest/d21/tables/dt21_322.10.asp
[4] A. Bandura, *Social Learning Theory*, Englewood Cliffs, NJ, USA, 1977.
[5] E. Baran, S. C. Bilici, and C. Mesutoglu, "Moving STEM beyond schools: Students' perceptions about an out-of-school STEM education program," *Int. J. Edu. Math., Sci. Technol.*, vol. 4, no. 1, p. 9, Jan. 2016.
[6] R. G. Barker and P. V. Gump, *Big School, Small School: High School Size and Student Behavior*, 1964.
[7] M. A. Busseri, L. Rose-Krasnor, T. Willoughby, and H. Chalmers, "A longitudinal examination of breadth and intensity of youth activity involvement and successful development.," *Develop. Psychol.*, vol. 42, no. 6, pp. 1313–1326, 2006.
[8] (2023). *Club List Example 1*. [Online]. Available: https://ghs.greenwichschools.org/student-life/clubs-and-activities
[9] (2023). *Club List Example 2*. [Online]. Available: https://www.newcaneyisd.org/site/Default.aspx?PageType=1&SiteID=24&ChannelID=212&DirectoryType=6
[10] N. Darling, L. L. Caldwell, and R. Smith, "Participation in school-based extracurricular activities and adolescent adjustment," *J. Leisure Res.*, vol. 37, no. 1, pp. 51–76, Mar. 2005.
[11] E. L. Deci and R. M. Ryan, "Self-determination theory," *Handbook Theories Social Psychol.*, vol. 1, no. 20, pp. 416–436, 2012.
[12] J. Dewey, *Interest and Effort in Education* (The Middle Works), vol. 7, 1913.
[13] J. S. Eccles, B. L. Barber, M. Stone, and J. Hunt, "Extracurricular activities and adolescent development," *J. Social Issues*, vol. 59, no. 4, pp. 865–889, Dec. 2003. [Online]. Available: Journal-of-Social-Issues-2003-Eccles-Extracurricular-Activities-and-Adolescent-Development.pdf
[14] J. S. Eccles and J. Templeton, "Chapter 4: Extracurricular and other after-school activities for youth," *Rev. Res. Educ.*, vol. 26, no. 1, pp. 113–180, Jan. 2002.
[15] A. F. Farb and J. L. Matjasko, "Recent advances in research on school-based extracurricular activities and adolescent development," *Develop. Rev.*, vol. 32, no. 1, pp. 1–48, Mar. 2012.
[16] A. F. Feldman and J. L. Matjasko, "The role of school-based extracurricular activities in adolescent development: A comprehensive review and future directions," *Rev. Educ. Res.*, vol. 75, no. 2, pp. 159–210, Jun. 2005.
[17] A. F. Feldman and J. L. Matjasko, "Profiles and portfolios of adolescent school-based extracurricular activity participation," *J. Adolescence*, vol. 30, no. 2, pp. 313–332, Apr. 2007.
[18] J. A. Fredricks and J. S. Eccles, "'Developmental benefits of extracurricular involvement: Do peer characteristics mediate the link between activities and youth outcomes?" *J. Youth Adolescence*, vol. 34, pp. 507–520, Dec. 2005.
[19] J. A. Fredricks and J. S. Eccles, "Extracurricular involvement and adolescent adjustment: Impact of duration, number of activities, and breadth of participation," *Appl. Develop. Sci.*, vol. 10, no. 3, pp. 132–146, Jun. 2006.
[20] M. Gardner, J. Roth, and J. Brooks-Gunn, "Adolescents' participation in organized activities and developmental success 2 and 8 years after high school: Do sponsorship, duration, and intensity matter?" *Develop. Psychol.*, vol. 44, no. 3, p. 814, 2008.
[21] R. Gilman, J. Meyers, and L. Perez, "Structured extracurricular activities among adolescents: Findings and implications for school psychologists," *Psychol. Schools*, vol. 41, no. 1, pp. 31–41, Jan. 2004.
[22] E. J. Haller, "High school size and student indiscipline: another aspect of the school consolidation issue?" *Educ. Eval. Policy Anal.*, vol. 14, no. 2, pp. 145–156, 1992.
[23] S. Han, R. Capraro, and M. M. Capraro, "How science, technology, engineering, and mathematics (STEM) project-based learning (PBL) affects high, middle, and low achievers differently: The impact of student factors on achievement," *Int. J. Sci. Math. Educ.*, vol. 13, no. 5, pp. 1089–1113, Oct. 2015.
[24] A. Holland and T. Andre, "Participation in extracurricular activities in secondary school: What is known, what needs to be known?" *Rev. Educ. Res.*, vol. 57, pp. 437–466, 1987.
[25] C. Jackson, M. J. Mohr-Schroeder, S. B. Bush, C. Maiorca, T. Roberts, C. Yost, and A. Fowler, "Equity-oriented conceptual framework for K-12 STEM literacy," *Int. J. STEM Educ.*, vol. 8, no. 1, pp. 1–16, Jun. 2021.
[26] J. Johnson, "Children, robotics, and education," *Artif. Life Robot.*, vol. 7, nos. 1–2, pp. 16–21, Mar. 2003.
[27] S. Jung and E.-S. Won, "Systematic review of research trends in robotics education for young children," *Sustainability*, vol. 10, no. 4, p. 905, Mar. 2018.
[28] E. J. Kleinert, "Effects of high school size on student activity participation," *Bull. Nat. Assoc. Secondary School Principals*, vol. 53, no. 335, pp. 34–46, Mar. 1969.
[29] R. Larson, D. M. Hansen, and G. B. Moneta, "Differing profiles of developmental experiences across types of organized youth activities.," *Develop. Psychol.*, vol. 42, no. 5, pp. 849–863, Jan. 2006.
[30] J. Lave, *Situated Learning: Legitimate Peripheral Participation*. Cambridge, U.K.: Cambridge Univ. Press, 1991.
[31] K. Leithwood and D. Jantzi, "A review of empirical evidence about school size effects: A policy perspective," *Rev. Educ. Res.*, vol. 79, no. 1, pp. 464–490, Mar. 2009.
[32] R. M. Lerner, J. V. Lerner, J. B. Almerigi, C. Theokas, E. Phelps, S. Gestsdottir, S. Naudeau, H. Jelicic, A. Alberts, L. Ma, L. M. Smith, D. L. Bobek, D. Richman-Raphael, I. Simpson, E. D. Christiansen, and A. von Eye, "Positive youth development, participation in community youth development programs, and community contributions of fifth-grade adolescents: Findings from the first wave of the 4-H study of positive youth development," *J. Early Adolescence*, vol. 25, no. 1, pp. 17–71, Feb. 2005.
[33] J. G. MacKinnon and H. White, "Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties," *J. Econometrics*, vol. 29, no. 3, pp. 305–325, Sep. 1985.
[34] J. L. Mahoney, A. L. Harris, and J. S. Eccles, "Organized activity participation, positive youth development, and the over-scheduling hypothesis," *Social Policy Rep.*, vol. 20, no. 4, pp. 1–32, Dec. 2006.
[35] J. L. Mahoney, R. W. Larson, and J. S. Eccles, *Organized Activities As Contexts of Development: Extracurricular Activities, After School and Community Programs*. London, U.K.: Psychology Press, 2005.
[36] H. Marsh and S. Kleitman, "Extracurricular school activities: The good, the bad, and the nonlinear," *Harvard Educ. Rev.*, vol. 72, no. 4, pp. 464–515, Dec. 2002.
[37] M. J. Mataric, "Robotics education for all ages," in *Proc. AAAI Spring Symp. Accessible, Hands AI Robot. Educ.*, 2004, pp. 22–24.
[38] B. R. McNeal Jr., "Participation in high school extracurricular activities: Investigating school effects," *Social Sci. Quart.*, vol. 80, no. 2, pp. 291–309, 1999.

[39] B. Means, H. Wang, V. Young, V. L. Peters, and S. J. Lynch, ''STEM-focused high schools as a strategy for enhancing readiness for postsecondary STEM programs,'' *J. Res. Sci. Teach.*, vol. 53, no. 5, pp. 709–736, May 2016.
[40] K. A. Moore, C. Trends, J. L. Hatcher, C. Trends, S. Vandivere, C. Trends, and B. V. Brown. (2000). *Children's Behavior and Well-Being: Findings From the National Survey of America's Families*. [Online]. Available: https://www.urban.org/sites/default/files/publication/62926/900845-Children-s-Behavior-and-Well-Being.PDF
[41] D. L. Morgan and D. F. Alwin, ''When less is more: School size and student social participation,'' *Social Psychol. Quart.*, vol. 43, no. 2, p. 241, Jun. 1980.
[42] NCES. (2023). *National Center for Education Statistics*. [Online]. Available: https://nces.ed.gov/ccd/elsi/tableGenerator.aspx
[43] S. C. Peck, R. W. Roeser, N. Zarrett, and J. S. Eccles, ''Exploring the roles of extracurricular activity quantity and quality in the educational resilience of vulnerable adolescents: Variable- and pattern-centered approaches,'' *J. Social Issues*, vol. 64, no. 1, pp. 135–156, Mar. 2008.
[44] J. Pennington, R. Socher, and C. Manning, ''Glove: Global vectors for word representation,'' in *Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP)*, 2014, pp. 1532–1543.
[45] D. B. Reeves, ''The learning leader/the extracurricular advantage,'' *Learning*, vol. 66, no. 1, pp. 86–87, 2008.
[46] A. Rosenfeld and N. Wise, *The Over-scheduled Child: Avoiding Hyper-parenting Trap*. New York, NY, USA: St. Martin's, 2010.
[47] D. R. Schaefer, S. D. Simpkins, A. E. Vest, and C. D. Price, ''The contribution of extracurricular activities to adolescent friendships: New insights through social network analysis.,'' *Develop. Psychol.*, vol. 47, no. 4, pp. 1141–1152, May 2011.
[48] P. Schoggen and M. Schoggen, ''Student voluntary participation and high school size,'' *J. Educ. Res.*, vol. 81, no. 5, pp. 288–293, May 1988.
[49] P.-S. Seow and G. Pan, ''A literature review of the impact of extracurricular activities participation on students' academic performance,'' *J. Educ. Bus.*, vol. 89, no. 7, pp. 361–366, Oct. 2014.
[50] (2023). *Statsmodels*. [Online]. Available: https://www.statsmodels.org/
[51] N. G. Stevens and G. L. Peltier, ''A review of research on small-school student participation in extracurricular activities,'' *J. Res. Rural Educ.*, vol. 10, no. 2, pp. 116–120, Jan. 1994.
[52] Syracuse. (2023). *Syracuse Academy of Science Charter School, NCES School ID 360010405549, List of Clubs*. [Online]. Available: https://www.sascs.org/students-parents/student-activities
[53] A. Vanmeter-Adams, C. L. Frankenfeld, J. Bases, V. Espina, and L. A. Liotta, ''Students who demonstrate strong talent and interest in STEM are initially attracted to STEM through extracurricular experiences,'' *CBE—Life Sci. Educ.*, vol. 13, no. 4, pp. 687–697, Dec. 2014.
[54] A. Wigfield and J. S. Eccles, ''Expectancy-value theory of achievement motivation,'' *Contemp. Educ. Psychol.*, vol. 25, no. 1, pp. 68–81, Jan. 2000.
[55] S. Rocker Yoel and Y. J. Dori, ''FIRST high-school students and FIRST graduates: STEM exposure and career choices,'' *IEEE Trans. Educ.*, vol. 65, no. 2, pp. 167–176, May 2022.
[56] J. F. Zaff, K. A. Moore, A. R. Papillo, and S. Williams, ''Implications of extracurricular activity participation during adolescence on positive outcomes,'' *J. Adolescent Res.*, vol. 18, no. 6, pp. 599–630, Nov. 2003.

AMY TANG is currently with the Lynbrook High School. She is an Activist and a Leader in extracurricular clubs. Her research interests include the impact of socioeconomic factors on equal opportunities in education and how extracurricular activities can be improved to enhance high school education.

YANG WANG received the Ph.D. degree in computer science from The University of Texas at Austin. He is currently an Associate Professor with the Computer Science and Engineering Department, The Ohio State University. His research interest includes the systems field. He is also actively involved in community outreach to promote STEM education in high schools.

CHUNQIANG TANG received the Ph.D. degree in computer science from the University of Rochester. He is currently a Research Scientist and a Senior Director of Meta Platforms, where he works on AI and systems co-design. He has received multiple best paper awards, including IEEE Micro Top Picks 2023 and 2024.