Journal of Education and Learning; Vol. 8, No. 4; 2019
ISSN 1927-5250 E-ISSN 1927-5269
Published by Canadian Center of Science and Education

Investigating the Role of School-Based Extracurricular Activity Participation in Adolescents' Learning Outcomes: A Propensity Score Method

Hsien-Yuan Hsu[1], KoFan Lee[2], John Bentley[3] & Sandra Acosta[4]

[1] Children's Learning Institute, University of Texas Health Science Center at Houston, USA

[2] School of Applied Sciences, University of Mississippi, USA

[3] School of Pharmacy, University of Mississippi, USA

[4] College of Education, Texas A&M University, USA

Correspondence: Hsien-Yuan Hsu, The University of Texas Health Science Center at Houston 7000 Fannin St., Suite 2373I, Houston, Texas 77030, USA. E-mail: Hsien-Yuan.Hsu@uth.tmc.edu

Received: March 20, 2019 Accepted: April 18, 2019 Online Published: June 13, 2019

doi:10.5539/jel.v8n4p8 URL: https://doi.org/10.5539/jel.v8n4p8

Abstract

The purpose of this study was to apply a propensity score method that could control for selection bias at both the student-level and school-level in an investigation of the causal effect linking participation in school-based extracurricular activity (SBEA) to adolescents' learning outcomes. The data for this study were drawn from the Education Longitudinal Study of 2002 (ELS: 2002) data set. The final sample comprised 12,247 10th graders; 6,026 (49.20%) were males. A propensity score method incorporating marginal mean weighting through stratification was implemented to analyze the data. Results showed that 10th graders who had proper intensity of participation in SBEA (6–15 hours a week) slightly outperformed peers who did not participate in SBEA on the performance of mathematics achievement in 12th grade. Regarding the link between SBEA participation and adolescents' long-term learning outcomes, results indicated 10th graders in 2002 with low to moderate levels of intensity (i.e., 1–15 hours) were more likely to achieve higher education credentials by the year 2012 when compared to non-participating peers.

Keywords: adolescent, mathematics achievement, multilevel data, propensity score method, school-based extracurricular activity

1. Introduction

A large body of contemporary studies have examined the impact of school-based extracurricular activity (SBEA) participation on adolescent development, including academic achievement, psychological adjustment, social skills, and successful transitions into early adulthood (Eccles, Barbaer, Stone, & Hunt, 2003; Feldman-Farb & Matjasko, 2005, 2012; Mahoney & Vest, 2012; Seow & Pan, 2014). According to two systematic reviews conducted by Feldman-Farb and Matjasko (2005, 2012), recent studies have extended the research line of SBEA by investigating the overscheduling hypothesis—an inquiry of optimal times or frequencies of adolescents' participation in extracurricular activities (Busseri, Rose-Krasnor, Willoughby, & Chalmers, 2006; Denault & Poulin, 2009a; Fredricks, 2012; Mahoney & Vest, 2012). The exploration of the relationship between SBEA participation intensity and positive outcomes can inform educators as to a proper school context for developing adolescents' capacities required for success. The current study aligns with this research trend by focusing on the role of SBEA participation intensity in adolescents' academic development, specifically learning outcomes.

As the body of literature continues to grow, however, findings linking SBEA to adolescents' positive outcomes have been mixed (Feldman-Farb & Matjasko, 2005, 2012). As highlighted by Feldman-Farb and Matjasko, one methodological flaw that can explain such disparities in findings is selection bias. Selection bias occurs when the characteristics of students who participate in SBEA are different from those of students who do not participate and these characteristics are related to outcome differences (Shadish, Cook, & Campbell, 2002). For example, students who participate in SBEA and students who do not participate can be different in terms of their personal characteristics (e.g., gender, ethnicity) and family characteristics (e.g., parents' highest level of education, family composition). Furthermore, they might be different with respect to schools where they belong to (e.g., school type, urbanicity). Failing to control for selection bias resulting from both student-related and school-related characteristics can lead to a spurious relationship between SBEA and learning outcomes, and therefore interferes with causal inference in SBEA research (Morgan & Winship, 2010).

To overcome this challenge in SBEA research, the propensity score method (PSM) approach is highly recommended. Simply speaking, PSM reduces multiple dimensional covariates (e.g., gender, school type) to a one-dimensional score called a propensity score. The propensity score is the conditional probability of assignment to a particular treatment (e.g., SBEA participation; Rosenbaum & Rubin, 1983). After propensity scores are estimated, researchers can use propensity scores in the data analysis to estimate the effect of SBEA participation on student outcomes (Guo & Fraser, 2015). However, our review of the literature found only a few studies applied PSM approach to control for selection bias (e.g., Zarrett et al., 2009). Apparently, SBEA researchers have largely ignored PSM as an approach for controlling selection bias resulting from student-related and school-related characteristics.

To address this vacuum in the literature, the current study aims to demonstrate how PSM can be used in SBEA studies. We used data from the Education Longitudinal Study of 2002 (ELS: 2002), a nationally-representative study, that allowed us to generalize our findings to U.S. 10th graders in 2002. Participation in SBEA was measured by asking participants the number of hours they spent on SBEA in a typical week (intensity). We posed the following question: For 10th graders in 2002, did more hours per week of SBEA participation in 10th grade lead to better mathematics academic achievement in 12th grade and increase the likelihood of earned a credential from their last/currently attended postsecondary institution 8 years after high school graduation (i.e., acquiring higher education credentials by 2012)? In this study, we applied one of the PSM approaches, namely marginal mean weighting through stratification (MMW-S) approach (Hong, 2010, 2012), for data analysis.

2. Method

2.1 Sample

The data for this study were drawn from Education Longitudinal Study of 2002 (ELS: 2002) public-use data. ELS: 2002 included a nationally representative sample of 10th graders in 2002. ELS: 2002 used a two-stage sampling design. For more detailed information of sampling design refer to Ingels et al.'s (2014) report. The original data contained 15,244 eligible base year participants. Because students' mathematics achievement in 12th grade was an outcome of interest, we excluded students who did not complete the mathematics assessment in 2004 (e.g., out of school, homeschooled, early graduate, or dropout). Regarding another outcome, acquisition of a higher education credential by 2012, we did not exclude samples with missing data from the analytical sample. Instead, we treated missing data as a category in the outcome when analyzing the data. Additionally, students who did not provide information of extracurricular activity intensity in the base year were eliminated. The final sample comprised 12,247 10th graders of which 49.20% were males and 60.06% were White. Demographic information for the final sample is presented in Table 1.

Table 1. Descriptive statistics of demographic characteristics for the sample

Variable	Frequency
Gender	
Male	49.20%
Female	50.80%
Ethnicity	
White	60.06%
African American	11.90%
Hispanic	12.91%
Others	15.13%
English is Student's Native Language	
Yes	84.45%
No	15.55%
Family Composition	
Two parents	78.46%
Single parent	20.74%
Others	0.80%
Parents' Highest Level of Education	
Did not finish high school	4.78%
Graduated from high school or GED	18.25%
Attended 2-year school, no degree	10.87%
Graduated from 2-year school	10.61%
Attended college, no 4-year degree	11.63%
Graduated from college	23.78%
Completed Master's degree or equivalent	12.75%
Completed PhD, MD, other advanced degree	7.34%
Total Family Income	
Lower than $25,001	18.16%
$25,001–$50,000	29.68%
$50,001–$75,000	21.64%
Higher than $75,000	30.52%

Note. n = 12,247.

2.2 Measures

The variables used in the analyses are presented in three categories as follows.

2.2.1 Outcome Variables

Two outcomes were selected to represent 10th graders short- and long-term learning achievement. The first outcome was students' mathematics achievement in the 12th grade. A mathematics assessment in ELS: 2002 comprised items in arithmetic, algebra, geometry, data/probability, and advanced topics. For our analyses, we used standardized scores to describe students' performance (variable name: F1TXMSTD), which had a mean = 51.18 and SD = 9.88.

Another outcome included in our analyses was the acquisition of a higher education credential by 2012. ELS: 2002 followed up targeted samples in 2012, 8 years after high school graduation (Ingels et al., 2014). We used a variable (F3PSLCRED) which indicated whether or not the respondent earned a credential from their last/currently attended postsecondary institution. Among 12,247 students, 41.10% received a higher education credential by 2012, 44.85% did not, and 14.05% were unable or refused to provide information on this item.

2.2.2 Treatment Variable

Tenth graders in 2002 were interviewed with an open-ended question: "In a typical week, how much time do you spend on school-sponsored extracurricular activities (for example, sports, and school clubs)?" In the public-use data set, ELS: 2002 coded students' responses into a variable (variable name: BYS42) with 22 categories (0 hours, 1 hour, 2 hours, …, 21 or more hours). Following Mahoney, Harris, and Eccles (2006), we created five-time categories with 5-hour increments ending with 16 or more hours. The reason we ended with 16 or more hours rather than 21 or more hours was due to a very small number of 10th graders (1.68%) spending 21 or more hours, which was consistent with Mahoney et al.'s (2006) findings. Accordingly, five levels of intensity were created in this study: dosage 0 = 0 hours (32.68%); dosage 1 = 1–5 hours a week (30.42%); dosage 2 = 6–10 hours a week (19.31%); dosage 3 = 11–15 hours a week (12.26%); and dosage 4 = 16 or more hours a week (5.34%).

2.2.3 Covariates

We identified covariates for propensity score estimates according to Feldman-Farb and Matjasko's (2005) review study and previous studies (Denault & Poulin, 2009b). Those covariates related to extracurricular activity participation included personal characteristics (e.g., gender, age/grade, ethnicity, native language, number of grades repeated, self-expectation, program studied, IEP, and perceived school safety) and family characteristics (e.g., income, parents' highest level of education, family composition, number of in-home siblings, home literacy resources, and parental-expectation). In addition, school characteristics associated with extracurricular activity intensity (e.g., school type, urbanicity, region, and enrollment) were also included as school-level covariates (see Appendix A). Regarding the missing data in covariates, multiple imputation is preferred but in practice, a single imputation of missing data suffices if the imputation values are stable across multiple imputation (Hong, 2012). Missing data imputation was implemented using the MI command in Stata.

2.3 Analytical Plan

This study applied Hong's (2010) PSM approach, namely *marginal mean weighting through stratification* (MMW-S). The computed marginal mean weights were then used as sampling weighs. A detailed theoretical rationale for MMW-S is provided by Hong (2010, 2012). The MMW-S approach was adopted in our study because of its feasibility even though the within-school sample size was small. Furthermore, the MMW-S approach could be applied to multilevel data (Hong, 2010). More importantly, this approach could be applied to treatment variables measured on a binary scale and on an ordinal or nominal scale (Hong, 2012). This section briefly reviews the general six-step procedure for applying MMW-S, implemented in the current study (Hong, 2012). The first five steps were implemented using *Stata* 13 and we adopted *HLM* 7.01 in the last step to estimate the treatment effect.

Step 1: Estimate the propensity score. The treatment variable—extracurricular activity intensity—was on an ordinal scale with five levels of dosage. Considering the hierarchal structure in our data, we applied a two-level ordinal logistic regression model with random intercept to estimate the logit score. With the assumption of systematic relationships between covariates and treatment dosage, estimated logit scores were presumed to be monotonic across the dosage levels (Hong, 2012). Following Hong's (2012) suggestion, we used the logit score of being assigned to the first dosage level as the propensity score for further analyses.

Step 2: Check common support. A common support was determined by the minimum of the maximum values (upper bound) and the maximum of the minimum values (lower bound) of a logit propensity score among all treatment groups. The samples outside the bound had no counterfactual information. Therefore, these samples should be excluded from the analytic sample (Hong, 2012). Additionally, a comparison between original full sample and analytic sample was conducted to confirm the generalization of causal inference (Hong, 2012).

Step 3: Stratify the sample on the estimated propensity score. We implemented stratification by dividing subjects into five equal strata (Cochrane, 1968; Rosenbaum & Rubin, 1984), increasing strata when necessary. Note increasing the numbers of strata can lead to a proportion of treated participants in the stratum that is too small, which will increase marginal mean weight volatility (see Step 4; Hong, 2012).

Step 4: Compute the marginal mean weight. As defined by Hong (2012), let n_s denote the sample size of stratum *s* and $n_{z=z,s}$ denote the number of subjects assigned to dosage level *z* in stratum *s*. The MMW-S for units assigned to treatment group *z* in stratum *s* were computed as

$$\frac{n_s \times \Pr(Z=z)}{n_{z=z,s}} \tag{1}$$

The numerator was the number of subjects in stratum *s* assigned to treatment group *z* in a completely randomized experiment. The marginal probability of assignment to treatment group *z* was $\Pr(Z = z)$, which was determined by the total proportion of participants assigned to group *z* in the observed data. The denominator was the number of subjects in stratum *s* actually assigned to treatment group *z* (Hong, 2012). If the denominator ($n_{z=z,s}$) < the numerator [$n_s \times \Pr(Z = z)$], the number of subjects assigned to dosage level *z* will be underrepresented in stratum *s*. A weight larger than 1.0 will be obtained to compensate for underrepresentation. Alternatively, if the denominator > the numerator, the number of participants assigned to dosage level *z* will be overrepresented in stratum *s* and a weight smaller than 1.0 will be obtained.

Step 5: Check balance. Weighted global tests were needed to verify the balance in propensity scores and covariates between multidosage treatment groups. Following Hong (2012), one-way ANOVAs incorporating the weights were conducted to test the mean difference in the estimated propensity score and covariates within each stratum. Considering the large sample size in our study, the results of F tests in ANOVA should not be the only criterion to verify balance. Therefore, we also computed η^2 (SOS_B/SOS_T) to show the ratio of between-group variation to total variation (i.e., the proportion of total variation accounted for). An η^2 close to zero indicated no between-group difference. Covariates remaining significantly different in multidosage treatment groups could not exceed 5%. Otherwise, researchers might need to modify the propensity score model (return to Step 1) or restratify the sample (return to Step 3; Hong, 2012).

Step 6: Analyze a weighted outcome model to estimate the treatment effect. A two-level model, incorporating MMW-S at Level-1, was applied for estimating the effect of extracurricular activity intensity on mathematics achievement in 12th grade, taking into account the dependency of subjects and cluster's impact (Hong, 2010; Hong & Raudenbush, 2006; Thoemmes & West, 2011):

Level-1 model: (2)

$$Y_{ij} = \beta_{0j} + \beta_{1j}Intensity1_{ij} + \beta_{2j}Intensity2_{ij} + \beta_{3j}Intensity3_{ij} + \beta_{4j}Intensity4_{ij} + r_{ij}.$$

Level-2 model: (3)

$$\beta_{0j} = \gamma_{00} + \mu_{oj},$$

$$\beta_{1j} = \gamma_{10} + \mu_{1j},$$

$$\beta_{2j} = \gamma_{20} + \mu_{2j},$$

$$\beta_{3j} = \gamma_{30} + \mu_{3j},$$

$$\beta_{4j} = \gamma_{40} + \mu_{4j}.$$

where $Intensity1_{ij}$ (dosage1: 1–5 hours), $Intensity2_{ij}$ (dosage2: 6–10 hours), $Intensity3_{ij}$ (dosage3: 11–15 hours), and $Intensity4_{ij}$ (dosage4: > 16 hours) are dummy indicators for four of the five multidosage treatment groups (reference group = dosage0: 0 hours). The effect of extracurricular activity intensity was assessed by the magnitude of the estimates of γ_{10}, γ_{20}, γ_{30}, and γ_{40}, which showed the expected differences on mathematics achievement between four dosage groups (dosage1 to dosage4) and the reference group (dosage0), respectively. Moreover, r_{ij} was a subject-specific random effect; μ_{1j}, μ_{2j}, μ_{3j}, and μ_{4j} were school-specific random effects (Hong, 2010).

Another outcome, acquisition of higher education credential, included three categories: receiving higher education credential by 2012 (code = 1), no credential by 2012 (reference group; code = 3), and no response (code = 2). Several authors do not recommend imputing no response category in the outcome (Allison, 2001; Cohen, Cohen, West, & Aiken, 2003). However, excluding the no response category from the analysis might lead to biased conclusions or limit the generalizability of study findings. Therefore, we used a two-level multinomial logistic regression model (i.e., baseline-category logits) with no credential by 2012 as the baseline category to estimate the effect of extracurricular activity intensity (Raudenbush & Bryk, 2002).

Since we had three categories in the outcome, two level-1 equations were specified. The first equation models—the likelihood of "receiving" vs. "not receiving" a higher education credential—were given dummy indicators for four of the five multidosage treatment groups (reference group = dosage0: 0 hours). These treatment groups were selected because they were our principal interest. The second equation modeled the likelihood of "no response" vs. "not receiving" a higher education credential. This model was beyond the scope of the present study and the results from this equation are not presented or discussed further.

Level-1 model: (4)

$$\eta_{1ij} = \beta_{0j(1)} + \beta_{1j(1)}Intensity1_{ij} + \beta_{2j(1)}Intensity2_{ij} + \beta_{3j(1)}Intensity3_{ij} + \beta_{4j(1)}Intensity4_{ij},$$

$$where\ \eta_{1ij} = ln\left(\frac{P\ (Y_{ij} = \text{"receiving"})}{P\ (Y_{ij} = \text{"not receiving"})}\right)$$

$$\eta_{2ij} = \beta_{0j(2)} + \beta_{1j(2)}Intensity1_{ij} + \beta_{2j(2)}Intensity2_{ij} + \beta_{3j(2)}Intensity3_{ij} + \beta_{4j(2)}Intensity4_{ij}.$$

$$where\ \eta_{2ij} = ln\left(\frac{P\ (Y_{ij} = \text{"no response"})}{P\ (Y_{ij} = \text{"not receiving"})}\right)$$

In the level-2 equations, the fixed- and random-effects were freely estimated:

Level-2 model: (5)

$$\beta_{0j(1)} = \gamma_{00(1)} + \mu_{oj(1)},$$

$\beta_{1j(1)} = \gamma_{10(1)} + \mu_{1j(1)}$,
$\beta_{2j(1)} = \gamma_{20(1)} + \mu_{2j(1)}$,
$\beta_{3j(1)} = \gamma_{30(1)} + \mu_{3j(1)}$,
$\beta_{4j(1)} = \gamma_{40(1)} + \mu_{4j(1)}$,
$\beta_{0j(2)} = \gamma_{00(2)} + \mu_{oj(2)}$,
$\beta_{1j(2)} = \gamma_{10(2)} + \mu_{1j(2)}$,
$\beta_{2j(2)} = \gamma_{20(2)} + \mu_{2j(2)}$,
$\beta_{3j(2)} = \gamma_{30(2)} + \mu_{3j(2)}$,
$\beta_{4j(2)} = \gamma_{40(2)} + \mu_{4j(2)}$,

The multinomial two-level model was implemented with the multinomial logit link function (Raudenbush & Bryk, 2002). The effect of extracurricular activity intensity was evaluated by the magnitude of the estimates of $\gamma_{10(1)}$, $\gamma_{20(1)}$, $\gamma_{30(1)}$, and $\gamma_{40(1)}$. Here γs showed the expected differences on the likelihood of receiving a higher education credential between four dosage groups (dosage1 to dosage4) and the reference group (dosage0), respectively.

3. Results

Using ELS 2002 data, we investigated the impact of 10th graders' extracurricular activity intensity on their mathematics academic achievement in 12th grade and the acquisition of a higher education credential by 2012. Results of step 1 showed the estimated logit propensity score ranged from -2.74 to 2.00 with M = -0.81 and SD = 0.77. We then implemented the MMW-S method for the two outcomes separately.

In the second step, the common support of propensity scores between treatment and control participants was determined by the minimum of the maximum values (upper bound: -2.39) and the maximum of the minimum values (lower bound: 1.05) of logit propensity scores among the five dosage groups. Only 177 of 12,247 students (1.45%) were not in the common support and were excluded from the analytic sample, leaving a final sample size of 12,070. The reduction of the analytical sample size was trivial and did not alter the population to which the causal inference could be generalized. Table 2 presents stratifying the analytic sample based on logit propensity scores (the results of step 3) and the weights computed for each dosage group in different strata (the results of step 4).

Table 2. Marginal mean weight through stratification (MMW-S) for multidosage of extracurricular activity intensity

	Dosage of Extracurricular Activity Intensity									
	0 (0 hours)		1 (1–5 hours)		2 (6–10 hours)		3 (11–15 hours)		4 (> 15 hours)	
Stratum	*n*	MMW-S	*n*	MMW-S	*n*	MMW-S	*n*	MMW-S	*n*	MMW-S
1	280	2.783	633	1.166	690	0.682	569	0.522	242	0.535
2	458	1.701	786	0.939	641	0.735	373	0.797	156	0.829
3	698	1.116	837	0.882	468	1.005	286	1.039	125	1.035
4	1,011	0.771	794	0.929	355	1.325	178	1.669	76	1.703
5	1,448	0.538	640	1.153	198	2.375	80	3.715	48	2.696
Total	3,895		3,690		2,352		1,486		647	

Note. In the second step, 177 of 12,247 students (1.45%) were not in the common support and excluded from the analytic sample, leaving a final sample size of 12,070.

Results of step 5 suggested five dosage groups showed significant differences in the logit propensity score before weighting, $F(4, 12065) = 657.50$, $p < .001$, $\eta^2 = 21.80\%$. After weighting, the between-group difference was approximately equal to 0, $F(4, 12065) = 657.50$, $p = .020$, $\eta^2 = 0.10\%$. A trivial η^2 suggested the between-group difference was close to 0 and the statistically significant of the F test might have been due to large sample size. The same results held for approximately 95% of the covariates. The weighting approach successfully balanced propensity scores and covariates between multidosage treatment groups.

3.1 Outcome: Mathematics Academic Achievement in 12th Grade

Results of step 6 are presented in left side of Table 3. We found that the effect of intensity1 (estimate = 0.38, SE = 0.23, *p* = .09) and intensity4 (estimate = 0.56, SE = 0.42, *p* = .19) were not statistically significant. Thus, 10th graders in either the lowest (1–5 hours a week) or highest intensity (16 or more hours a week) groups performed similarly as students who spent 0 hours in SBEA in terms of their mathematics achievement scores. Alternatively, the mathematics achievement scores of 10th graders who spent 6–10 hours on extracurricular activities were higher by 1.48 points (intensity2, SE = 0.25, $p < .05$) than students who spent 0 hours. Similarly, 10th graders who spent 11–15 hours on extracurricular activity outperformed by 1.77 points (intensity3, SE = 0.31, $p < .05$) students who did not participate in extracurricular activities.

Table 3. Effects of the extracurricular activity intensity on mathematics achievement and acquisition of higher education credential by 2012

Outcome: Acquisition of Higher Education Credential			Outcome: Acquisition of Higher Education Credential			
Fixed Effects	Coefficient estimate (*SE*)	95% CI	Fixed Effects	Coefficient estimate (*SE*)	OR	95% CI for OR
Intercept, γ_{00}	50.253* (0.22)	(49.816, 50.691)	Intercept, $\gamma_{00(1)}$	-0.34* (0.05)	0.71	(0.646, 0.779)
Intensity1, γ_{10}	0.38 (0.23)	(-0.059, 0.828)	Intensity1, $\gamma_{10(1)}$	0.29* (0.06)	1.33	(1.192, 1.494)
Intensity2, γ_{20}	1.48* (0.25)	(0.984, 1.976)	Intensity2, $\gamma_{20(1)}$	0.32* (0.07)	1.38	(1.206, 1.577)
Intensity3, γ_{30}	1.77* (0.31)	(1.161, 2.375)	Intensity3, $\gamma_{30(1)}$	0.42* (0.10)	1.52	(1.279, 1.803)
Intensity4, γ_{40}	0.56 (0.42)	(-0.267, 1.378)	Intensity4, $\gamma_{40(1)}$	0.17 (0.12)	1.19	(0.958, 1,475)

Note. Intensity1 = one to five hours a week; Intensity2 = six to ten hours a week; Intensity3 = eleven to fifteen hours a week; and Intensity4 = sixteen hours or more a week. The reference group was students with 0 hours of participation. OR = odds ratio. *$p < .05$.

3.2 Outcome: Acquisition of Higher Education Credential by the Year 2012

Results of step 6 (estimated coefficients and odds ratios) are presented in right side of Table 3. We found that the effect of intensity1 (estimate = 0.29, *SE* = 0.06), intensity2 (estimate = 0.32, *SE* = 0.07), and intensity3 (estimate = 0.42, *SE* = 0.10) were statistically significant ($p < .05$). Generally speaking, compared to students who spent 0 hours, those who spent 1–5 hours, 6–10 hours, or 11–15 hours a week in SBEA, were more likely to receive a higher education credential by 2012. We interpreted the results with odds ratios rather than logits to better understand our results. The odds ratio associated with intensity1 was 1.33, suggesting that the odds of receiving a higher education credential for students who spent 1–5 hours a week in SBEA in 10th grade were 1.33 times the odds for students who spent 0 hours (i.e., the odds were 33% higher). The magnitude of odds ratio for intensity2 was 1.38, suggesting that the odds of receiving a higher education credential for students who spent 6–10 hours a week in SBEA in 10th grade were 1.38 times the odds for students who spent 0 hours (the odds were 38% higher). A relatively larger odds ratio (1.52) was found for intensity3, suggesting that the odds of receiving a higher education credential for students who spent 11–15 hours a week in extracurricular activities were 1.52 times the odds for students who spent 0 hours (the odds were 52% higher). Finally, the effect of intensity4 (estimate = 0.17, *SE* = 0.12, $p = .12$) was not statistically significant. Hence, there was no evidence supporting a difference in the odds of receiving a higher education credential between the highest intensity groups (16 or more hours) and the group that did not participate in extracurricular activities.

4. Discussion and Conclusion

The current study utilized one of PSM approaches, namely marginal mean weighting through stratification (MMW-S) method (Hong, 2010, 2012), to investigate the causal effect linking SBEA participation in 10th grade to short-term (mathematics achievement in 12th grade) and long-term (acquisition of a higher education credential by 2012) outcomes. Following Hong (2012), we implemented the MMW-S approach to ELS: 2002 data. Our results showed the MMW-S approach reasonably balanced the covariates between multidosage treatment groups. Therefore, the impact of SBEA participation on adolescents' learning outcomes can be viewed as a causal effect because the select bias was appropriately controlled.

The results show that 10th graders who had proper intensity of participation in SBEA (6–15 hours a week) outperformed peers who do not participate in SBEA on the performance of mathematics achievement in 12th grade (the estimated differences ranged from 1.48 to 1.77). Conversely, for those who had low intensity participation (1–5 hours a week), their mathematics achievement was not statistically different from non-participants. Likewise, 10th graders with high intensity participation (16 or more hours a week) performed similarly as non-participants on mathematics achievement. These results suggest a non-linear impact of intensity on adolescents' short-term learning outcomes (i.e., mathematics achievement) in the 12th grade. Our findings are consistent with Fredricks (2012), who also analyzed ELS: 2002 data and found higher intensity in SBEA led to the decline of short-term mathematics achievement.

However, our study provided a distinct finding inconsistent with previous studies. Fredricks (2012) found that even though the impact of intensity on mathematics achievement could be described as an inverted U shape, 10th graders with highest intensity in SBEA (16 or more hours a week) still outperformed non-participants on mathematics achievement, and Mahoney et al. (2006) had a similar conclusion. Our result did not support their findings. We found no statistical difference on mathematics performance between those highest-intensity participants and non-participants. That is, the short-term academic benefits to SBEA were limited to moderate intensity of participation (6–15 hours a week). Furthermore, the findings in terms of this short-term learning outcome need to be interpreted considering practical significance (Thompson, 2006). The magnitudes of statistically significant coefficient estimate in Table 3 ranged from 1.48 to 1.77, which were relatively small compared with the SD (9.88) for mathematics achievement.

Regarding the link between SBEA participation and adolescents' long-term learning outcomes, the results indicated 10th graders with low to moderate levels of intensity (i.e., 1–15 hours) were more likely to achieve higher education credentials by 2012 (odds ratios ranged from 1.33 to 1.52 in Table 3), compared with non-participants. However, there was no statistical difference between 10th graders with the highest intensity participation (16 or more hours a week) and non-participation. Our results suggest that there is a non-linear effect of SBEA participation on long-term educational attainment. This non-linear relationship was also supported by Fredricks' (2012) findings, where Fredricks used adolescents' educational status two years after high school (e.g., high school diploma, enrolled in 2-year/4-year college or university) as a long-term learning outcome.

Despite this study's potential contributions to the literature, three limitations are noteworthy. First, the current study only considered academic success of adolescents as outcomes. Future studies can apply MMW-S approach on different nationally representative data to determine the causal effect linking SBEA participation to adolescents' risky behaviors, internalizing problems, civic development (e.g., Denault & Poulin, 2009a), well-being, and interpersonal functioning (Busseri et al., 2006). Second, our study applied a quantitative approach which is not capable of fully describing the mechanism behind the causal effect linking SBEA participation to learning outcomes. We suggest future studies to qualitatively explore the factors which mediate the causal effect, such as perseverance, time management and autonomous acts, in order to better depict the experience of youth academic success. Future quantitative studies can subsequently test such mediated relationships. Last, this study investigated one dimension of SBEA participation. Other dimensions of SBEA participation, such as breadth (total number of activities) and duration (length of participation overtime), could be considered in the future studies (Feldman-Farb & Matjasko, 2012).

In conclusion, moderate intensity of SBEA participation (6–15 hours a week) can benefit adolescents' short- and long-term learning outcomes. However, these academic benefits from SBEA were not observed at the highest level of intensity (16 or more hours a week). Therefore, parents and teachers are encouraged to consider SBEA as a means to promote adolescents' academic success but should pay attention to those highly engaged adolescents. Moreover, we also conclude that although the lowest intensity (i.e., 1–5 hours a week) did not lead to outperformance in short-term learning outcome, it boosted the likelihood of achieving higher education credentials. For this reason, parents and teachers should provide non-participating adolescents more opportunities to increase their participation. Even a few hours per week might make a difference with respect to adolescents' long-term academic success.

References

Allison, P. D. (2001). *Missing data*. Thousand Oaks, CA: Sage.

Busseri, M. A., Rose-Krasnor, L., Willoughby, T., & Chalmers, H. (2006). A longitudinal examination of breadth and intensity of youth activity involvement and successful development. *Developmental Psychology*, *42*(6), 1313–1326. https://doi.org/10.1037/0012-1649.42.6.1313

Cochrane, W. G. (1968). The effectiveness of adjustment by subclassification in removing bias in observational studies. *Biometrics, 24*(2), 295-313. https://doi.org/10.2307/2528036

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: LEA.

Denault, A.-S., & Poulin, F. (2009a). Intensity and breadth of participation in organized activities during the adolescent years: Multiple associations with youth outcomes. *Journal of Youth and Adolescence*, *38*(9). https://doi.org/10.1007/s10964-009-9437-5

Denault, A.-S., & Poulin, F. (2009b). Predictors of adolescent participation in organized activities: A five-year longitudinal study. *Journal of Research on Adolescence*, *19*(2), 287–311.

https://doi.org/10.1111/j.1532-7795.2009.00597.x

Eccles, J. S., Barbaer, B. L., Stone, M., & Hunt, J. (2003). Extracurricular activities and adolescent development. *Journal of Social Issues*, *59*(4), 865–889. https://doi.org/10.1046/j.0022-4537.2003.00095.x

Feldman-Farb, A., & Matjasko, J. L. (2005). The role of school-based extracurricular activities in adolescent development: A comprehensive review and future directions. *Review of Educational Research*, *75*(2), 159–210. https://doi.org/10.3102/00346543075002159

Feldman-Farb, A., & Matjasko, J. L. (2012). Recent advances in research on school-based extracurricular activities and adolescent development. *Developmental Review, 32*(1), 1–48. https://doi.org/10.1016/j.dr.2011.10.001

Fredricks, J. A. (2012). Extracurricular participation and acaemic outcomes: Test the over-scheduling hypothesis. *Journal of Adolescence*, *41*, 295–306. https://doi.org/10.1007/s10964-011-9704-0

Guo, S., & Fraser, M. W. (2015). *Propensity score analysis: Statistical methods and applications* (2nd ed.). Thousand Oaks, CA: Sage.

Hong, G. (2010). Marginal mean weighting through stratification: Adjustment for selection bias in multilevel data. *Journal of Educational and Behavioral Statistics*, *35*(5), 499–531. https://doi.org/10.3102/1076998609359785

Hong, G. (2012). Marginal mean weighting through stratification: A generalized method for evaluating multivalued and multiple treatments with nonexperimental data. *Psychological Methods*, *17*(1), 44–60. https://doi.org/10.3102/1076998609359785

Hong, G., & Raudenbush, S. W. (2006). Evaluating kindergarten retention policy: A case study of causal inference for multilevel observational data. *Journal of the American Statistical Association, 101*(475), 901–910. https://doi.org/10.1198/016214506000000447

Ingels, S. J., Pratt, D. J., Alexander, C. P., Jewell, D. M., Lauff, E., Mattox, T. L., & Wilson, D. (2014). *Education Longitudinal Study of 2002 third follow-up data file documentation (NCES 2014-364)*. Retrieved from Washington, DC: http://nces.ed.gov/pubs2014/2014364.pdf

Mahoney, J. L., Harris, A. L., & Eccles, J. S. (2006). Organized activity participation, positive youth development, and the over-scheduling hypothesis. *Social Policy Report*, *20*(4), 3–31. https://doi.org/10.1002/j.2379-3988.2006.tb00049.x

Mahoney, J. L., & Vest, A. E. (2012). The over‐scheduling hypothesis revisited: Intensity of organized activity participation during adolescent and young adult outcomes. *Journal of Research on Adolescent*, *22*(3), 409–418. https://doi.org/10.1111/j.1532-7795.2012.00808.x

Morgan, S. L., & Winship, C. (2010). *Counterfactuals and causal inference: Methods and principles for social research*. New York, NY: Cambridge University Press.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods* (2nd ed.). Thousand Oaks, CA: Sage.

Rosenbaum, P. R., & Rubin, D. B. (1984). Reducing bias in observational studies using subclassification on the propensity score. *Journal of American Statistical Association*, *79*(387), 516–524. https://doi.org/10.1080/01621459.1984.10478078

Seow, P.-S., & Pan, G. (2014). A literature review of the impact of extracurricular activities participation on students' academic performance. *Journal of Education for Business*, *89*(7), 361–366. https://doi.org/10.1080/08832323.2014.912195

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Belmont, CA: Wadsworth.

Thoemmes, F., J,, & West, S. G. (2011). The use of propensity scores for nonrandomized designs with clustered data. *Multivariate Behavioral Research*, *46*, 514–543. https://doi.org/10.1080/00273171.2011.569395

Thompson, B. (2006). *Foundations of Behavioral Statistics: An Insight-based Approach*. New York, NY: Guilford Press.

Zarrett, N., Fay, K., Li, Y., Carrano, J., Phelps, E., & Lerner, R. M. (2009). More than child's play: Variable- and pattern-centered approaches for examining effects of sports participation on youth development. *Developmental Psychology*, *45*(2), 368–382. https://doi.org/10.1037/a0014577

Appendix A

Covariates for Propensity Score Estimates

Variable Name in ELS 2002	Description
Outcome variables	
F1TXMSTD	F1 math standardized score
F3PSLCRED	Whether respondents earned a credential from their last/currently attended post-secondary institution
Treatment variable	
BYS42	Hours/week spent on extracurricular activities
Covariates (student-level)	
BYSTUWT	Base year student weight
BYSEX	Sex-composite
BYRACE	Student's race/ethnicity-composite
BYSTLANG	Whether English is student's native language-composite
BYDOB_P	Student's year and month of birth
BYFCOMP	Family composition
BYSIBHOM	BY number of in-home siblings
BYPARED	Parents' highest level of education
BYINCOME	Total family income from all sources 2001-composite
BYGRDRPT	Number of grades repeated (K-10)
BYSTEXP	How far in school student thinks will get-composite
BYPARASP	How far in school parent wants 10th grader to go-composite
BYSCHPRG	High school program reported by student-composite
BYIEPFLG	Base year Individualized Education Plan
BYWORKSY	Student held job for pay during 2001-2002 school year
BYSCSAF2	BY school safety index: student's perceptions
Covariates (school-level)	
BYSCTRL	School control
BYURBAN	School urbanicity
BYREGION	Geographic region of school
BYG10EP	Grade 10 enrollment-2001/02 school roster-categorical