Pre-Analysis Plan
Impact Evaluation of Boys and Girls Club Academic and Emotional Well-Being Enrichment Programs

Michael Kofoed
University of Tennessee

Jonathan Tebes
University of Notre Dame

July 2024

I. Overview

This is a pre-analysis plan for the randomized evaluation of the Boys and Girls Club academic and emotional well-being enrichment program. We begin with the motivation for the program and the randomized evaluation in Section II. Section III lays out the target population, recruitment strategies, and eligibility criteria. Section IV describes our planned data sources, and Section V defines our primary and secondary outcomes. In Section VI, we present and discuss our power calculations relative to observational evidence and related studies. Finally, our empirical strategy is laid out in Section VII.

II. Background

The healthy development of social-emotional skills, including behavioral and emotional regulation, leadership, decision-making, self-esteem, and relationship-building, is important for children's present and future well-being and success. For instance, behavioral and emotional self-regulation allow youth to healthily relate to others in the present, promote healthy social functioning in adulthood, and explore their own interests and aspirations, thus developing motivations for future careers (Pulkkinen et al., 2002, Oliveira et al, 2015). Despite this, many youth do not receive adequate emotional support and experience a decreased sense of belonging in school (Borman et. al, 2019). Chronic absenteeism remains at an all time high (Mervosh and Parish, 2024) and school enrollment has failed to recover to pre-pandemic levels (Burtis et al., 2024).

A range of youth programs that support the development of social-emotional skills have been studied previously, displaying varying degrees of effectiveness. Some youth programs that incorporate social-emotional learning (Battistich et. al, 2004; Payton et al., 2008; Zimmerman et. al, 2018) and trauma-informed therapy (Bhatt et al., 2023) have been found to improve educational outcomes for disadvantaged populations, while in-school SEL interventions were found to have no consistent effect (Crean and Johnson, 2013, Berry et. al, 2016).

The Boys & Girls Club of the Northern Indiana Corridor (BGCNIC) offers one potential solution catered toward the needs of elementary school children in the South Bend and Mishawaka communities, incorporating elements of social-emotional learning (SEL) and academic tutoring. A study of their after-school program will contribute further evidence on the importance of SEL programming, providing critical information on how SEL may complement academic programming, especially for children who struggle with behavioral problems.

III. Evaluation Design

A. Research Questions

Does participation in the academic and emotional well-being enrichment programs impact academic outcomes including grades, standardized test scores, and school attendance as well as behavioral outcomes measured through incident reports and teacher-reported behavior?

B. Eligibility

Club members are only eligible for academic and emotional well-being enrichment programs if they are entering kindergarten, first, second, third, fourth, or fifth grade in the year they enroll in the study and if they apply to attend one of the participating club sites. There are 10 of these club sites within the partner school districts in St. Joseph County. These districts include South Bend Empowerment Zone (SBEZ), School City of Mishawaka (SCM), South Bend Community School District (SBCSD), and South Bend charter schools.

C. Randomization

In early fall, the research team will randomize study participants into three groups: control, SEL, and SEL-plus. The control group will continue with standard club programming. In addition to standard club programming, the SEL group will receive two 45-minute SEL-focused programming sessions. The SEL-plus group will receive standard club programming, two 45-minute SEL-focused programming sessions, and two 45-minute academic-focused programming sessions (one STEM and one literacy).

We plan to enroll students into the study for three years, with a total of about 1,110 study participants. We estimate that 370 students will be enrolled in the two treatment groups, and 370 students will be enrolled in the control group over the course of the study. These numbers may vary depending on the take-up rate for academic and emotional well-being enrichment program participation; for example, if a higher rate of students originally assigned to the treatment group turn down the program, more students will need to be randomized into treatment in subsequent years in order to fill the spots that become available in academic and emotional well-being enrichment program programming.

D. Intervention

The club enrichment program intervention is an expansion of club programming designed to recover some of the learning loss of the COVID-19 pandemic, especially for disadvantaged students. The year-long enrichment programs are designed to use the innate passions of children to empower them to learn outside of the classroom during their time at club each day. All club activities are directed by youth development professionals (YDPs). The control group will participate in standard club programming, which consists of snack, homework time, access to therapy services, access to

Beable (an online literacy-development platform), and high-yield activities including leisure reading, writing activities and games like chess or Scrabble that develop cognitive skills. The EWB treatment will have access to these same standard club program elements, but they will also attend two 45-minute sessions on SEL. The SEL plus academic treatment group will attend two 45-minute SEL sessions, one 45-minute literacy session, and one 45-minute STEM-focused session. The lesson plans for these sessions are collaboratively developed with content-expert community partners at the Robinson Learning Center and Riverbend Community Math Center.

IV. Data Sources

A. Indiana Department of Education (IDOE)

Through a data sharing partnership with IDOE, the research team will pull academic outcomes for the students as well as other administrative data.

B. School districts in St. Joseph County

LEO is pursuing data sharing agreements with all school districts in St. Joseph County that are involved in the study. Through this partnership, the research team can pull academic outcomes of interest like grades, attendance rate, and behavioral incidents. The research team will also administer the strengths and difficulties questionnaire (SDQ) to the teachers of study participants.

C. Boys and Girls Club of St. Joseph County

BGCNIC gathers baseline information on new program participants that they store internally. Program records from BGCNIC will provide engagement measures including enrollee attendance to various events, meetings, and social worker interactions which will allow for descriptive work of the treatment and treatment on the treated analysis. LEO has a data sharing agreement with BGCNIC.

V. Study Outcomes

BGCNIC's academic and emotional well-being enrichment program is designed to improve academic and behavioral/emotional outcomes for young students. A key component of this program is the provision of a safe space for young children to learn and develop self-regulatory behaviors that may prove to have an impact on their academic outcomes. That is why this evaluation will primarily focus on the educational outcomes of participants, pulled from administrative data in the partnering school districts. One-year outcomes will be collected and reported for the first enrollment cohort in summer 2025, and preliminary results will be updated and reported for each subsequent year of the study. Three-year results will be available for the first enrollment cohort in summer of 2027, and for the full sample in the summer of 2029.

A. Primary Outcomes

- Standardized GPA: cumulative GPA at the end of each school year (when available), normalized by the control group mean and standard deviation at the end of year 1 within a given boys and girls club site.

- Standardized index consisting of standardized test scores in I-ready math, I-ready ELA and HMH reading assessments. Each standardized test will receive equal weight and be normalized using the mean and standard deviation of the control group in year 1 at a given BGCNIC program site.
    - Math (I-ready assessment), normalized using the mean and standard deviation of the control group in year 1 at a given BGCNIC program site
    - ELA (I-ready assessment), normalized using the mean and standard deviation of the control group in year 1 at a given BGCNIC program site.
    - Reading (MHM assessment), normalized using the mean and standard deviation of the control group in year 1 at a given BGCNIC program site.

  We will also report average treatment effects for each component of this index separately as secondary outcomes.
- Social emotional skills: We will construct a social-emotional (non-cognitive) skill index gleaned from administrative records of student behavior. Depending upon the availability of school record data, we may follow Rose et al. (2022). For each year, such an index would include the following measures:
    - Disciplinary outcomes (suspensions, detentions, reported incidents, time outs, etc.)
    - Attendance rate (fraction of school days attended)
    - Indicator for whether participant has been chronically absent
    - Grade repetition

  We will also report average treatment effects for each component of this index separately as secondary outcomes.

B. *Secondary Outcomes*

- Chronic absenteeism: indicator for whether participant has been chronically absent
- School attendance rate: fraction of school days attended
- Grade repetition: indicator for whether the student did not progress to the next grade on time
- Homework completion: cumulative rate of homework completion as reported by school records
- Disaggregated standardized test scores: Math I-ready, ELA I-ready, Reading MHM assessment
- Behavioral incidents: total number of suspensions, detentions, and other reported incidents, as reported in school disciplinary records.
- Classroom behavior: total score from the teacher-facing Strengths and Difficulties Questionnaire. We also may report sub-categories of outcomes to explore mechanisms.

VI. Statistical Power and Sample Size

We plan to enroll students into the study for three years, with a total of about 1,110 study participants and a program take-up rate of 90%. We estimate that 370 students will be enrolled in the two treatment groups, and 370 students will be enrolled in the control group over the course of the study. We are powered to detect a 0.1845 standard deviation change in grades and standardized test scores between the control group and either treatment group, or between the two treatment groups.

VII. Empirical Strategy

A. Main Specification

We will estimate intent-to-treat (ITT) treatment effects by OLS using the following regression:

$$(1) \qquad Y_i = \theta_1 T_{1i} + \theta_2 T_{2i} + X_{ist}\theta_3 + \mu_{1s} + \lambda_{1t} + \epsilon_{1ist}$$

where Y_i is the outcome. T_{1i} is an intent-to-treat dummy indicating the random assignment of person *i* to the SEL group, and T_{2i} is an intent-to-treat dummy indicating the random assignment of person *i* to the SEL + academic arm of the experiment. The vector X_{ist} includes a set of person-level characteristics collected at baseline. We will also control for club site fixed effects (μ_{1s}) and year fixed effects (λ_{1t}), and ϵ_{1ist} is an error term. The coefficient on the treatment dummies, θ_1 and θ_2 will give us the difference in means between the treatment and comparison groups for each treatment group, or the estimated impact of the program. We will also test the null hypothesis that $\theta_1 = \theta_2$ or that the two treatment arms have the same impact.

B. Treatment on Treated Specifications

In addition to the reduced-form estimates obtained in the equations above, we are also interested in estimating the causal impact of enrichment program participation, also known as the *treatment-on-treated* (TOT) effect. In this case, not all club members assigned to enrichment program treatment arms will take up programming sessions. To this end, we will estimate the TOT by instrumenting for intervention participation with treatment assignment using a two-stage least squares regression framework.

There are two possible models. Using only those assigned to the SEL group (added SEL sessions only) and the control group, we can examine the impact of the receipt of SEL sessions on outcomes. In this case, let S_i be a dummy that equals 1 if the person participated in at least one SEL session. The equation of interest in this case can be described by the equation

$$(2) \qquad Y_i = \beta_0 + \beta_1 S_i + X_{ist}\beta_3 + \mu_{2s} + \lambda_{2t} + \epsilon_{2ist}$$

As S_i is endogenous, we would need to use the assignment to the SEL treatment group where the first-stage regression is then

$$(3) \qquad S_i = \Phi_0 + \Phi_1 T_{1i} + X_{ist}\Phi_3 + \mu_{3s} + \lambda_{3t} + \epsilon_{3ist}$$

We can so examine whether the dosage of SEL programming matters, where S_i can be replaced with the percentage of total SEL sessions attended. We will also estimate this TOT model for the second treatment arm where we instrument for SEL + Academic attendance with random assignment into the SEL + academic study group.

In all models, we will use heteroskedasticity-robust standard errors. There is some between-year variation in club site capacity and participant engagement, so standard errors will be clustered at the levels of the club site by year throughout the study time period.

C. Treatment Effect Heterogeneity and Subgroup Analyses

Given that this study will recruit a broad range of students, the enrichment programs will likely have different effects within different sub-groups. Understanding whether the enrichment programs work broadly, for some sub-population of policy interest, or is most beneficial for some surprising sub-sample provides crucial information to Boys & Girls Club of America, school districts, and governments on how they might scale this program in the event of a positive finding.

The study will estimate the impact of enrichment programs across several outcome categories and subgroups. The research team is interested in determining whether the intervention is more effective for certain populations relative to others. Our main sub-group analysis will explore heterogeneity of treatment effects by baseline social-emotional well-being indices measures prior to randomization. In particular, we are interested in understanding whether additional social-emotional support is necessary for students with behavioral difficulties before they can achieve academically.[1] A trained psychologist on staff at BGCNIC collects EWB screeners according to the parent-report Diagnostic and Statistical Manual of Mental Disorder (DSM-5) cross-cutting measures on depression, anxiety, and anger. We will split the sample at the median of this index but will also graph treatment effects across the distribution of the index.

Primary subgroup

- Baseline social-emotional well-being index: consisting of standardized T-Score categories (none to slight, mild, moderate, severe) based upon parent-report Diagnostic and Statistical Manual of Mental Disorder (DSM) cross-cutting measures for depression, anxiety, and anger.

We are also interested in exploring heterogeneity by their baseline cognitive skills index, gender, and race. To construct the cognitive skills index, we will use grades, standardized test scores, and other academic information prior to randomization to predict baseline cognitive measures.[2]

Other Primary Subgroups

- Baseline cognitive measures index: we will combine grades and standardized test scores prior to randomization into an index to capture baseline cognitive/academic performance.
- Student gender

[1] In the extreme case, these students may face a Leontief educational production function.

[2] For example, if these predictions have a low R2, we may instead explore heterogeneity by baseline grades and attendance rates.

- Student race

Secondary Subgroups

- Baseline non-cognitive measures index: we will combine disciplinary outcomes, attendance, and grade repetition (and homework completion if available) prior to randomization to measure non-cognitive skills at baseline.
- To measure level of poverty and family resources at baseline: consisting of household income, household structure, benefits usage (SNAP/TANF). We will split the sample at the median of the baseline index for poverty level and family resources.
- Student academic grade (at point of randomization)

D. Exploring Heterogeneity using Machine Learning

We may draw on an emerging literature that leverages machine-learning methods to explore heterogeneity of causal effects (Chernozhukov et al., 2018; Athey and Imbens, 2015, 2019; Davis and Heller, 2017). This methodology will enable us to learn as much as possible from our data using a disciplined and data-driven approach. Since the "state of the art" is still evolving, we cannot commit to a particular approach at present. However, we plan to pre-specify our approach prior to running this analysis and will interpret our results as suggestive.

E. Multiple Hypothesis Testing

Testing multiple hypotheses raises the likelihood that any one hypothesis is found to be statistically significant purely by chance. We will supplement our results by reporting summary indices that aggregate multiple outcome variables within a common outcome domain. Aggregation not only improves the statistical power within a given domain but also vastly reduces the number of hypotheses examined. This plan pre-specifies what data will be collected, primary and secondary outcomes, the main specification, and subgroups of interest. By committing to a set of analyses in advance, we avoid concerns about data-mining and specification searching, and credibly commit to a few hypotheses that, together, comprise the central test of BGCNIC's model. Classic p-values will be reported for all outcomes, which will provide a reader with full information that they can use to make multiple hypothesis testing corrections if they desire. We will also conduct non-parametric permutation tests and report permuted p-values for the main sets of analyses following Chetty et al. (2016).[3]

[3] This approach entails randomly re-assigning treatment status to students in the main sample and running the main specification thousands of times to simulate a counterfactual distribution of T-statistics. Relative to this counterfactual distribution, we can then compute permuted p-values as likelihood of observing our realized T-statistic. The same approach can be applied to sets of hypotheses to calculate the likelihood of observing by chance the magnitudes of treatment effects observed in the study.

VIII. References

Battistich, V., Schaps, E. & Wilson, N. Effects of an Elementary School Intervention on Students' "Connectedness" to School and Social Adjustment During Middle School. The Journal of Primary Prevention 24, 243–262 (2004).

Berry, V., Axford, N., Blower, S. et al. The Effectiveness and Micro-costing Analysis of a Universal, School-Based, Social–Emotional Learning Programme in the UK: A Cluster-Randomised Controlled Trial. School Mental Health 8, 238–256 (2016). https://doi.org/10.1007/s12310-015-9160-1

Bhatt, Monica P. Randomized evaluation of a school-based, trauma-informed group intervention for young women in Chicago. ScienceAdvances (2023). DOI:10.1126/sciadv.abq2077.

Borman, G. D., Rozek, C. S., Pyne, J., & Hanselman, P. (2019). Reappraising academic and social adversity improves middle school students' academic achievement, behavior, and well-being. Proceedings of the National Academy of Sciences, 116(33), 16286–16291.

Burtis, E., Sofoklis Goulas, et al. (2024, March). Breaking Down Enrollment Declines in Public Schools. Brookings.

Chetty, Raj et al. "The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment." *The American economic review* vol. 106,4 (2016): 855-902. doi:10.1257/aer.20150572

Crean, H. F., & Johnson, D. B. (2013). Promoting Alternative Thinking Strategies (PATHS) and elementary school aged children's aggression: results from a cluster randomized trial. American journal of community psychology, 52(1-2), 56–72. https://doi.org/10.1007/s10464-013-9576-4

Mervosh, Sarah, and Francesca Paris. (2024, March 29). Why School Absences Have 'Exploded' Almost Everywhere. The New York Times.

Oliveira, I. M., do Ceu Taveira, M., & Porfeli, E. J. (2015). Emotional aspects of childhood career development: importance and future agenda. International Journal for Educational and Vocational Guidance, 15, 163–174. https://doi-org.proxy.library.nd.edu/10.1007/s10775-015-9303-9

Payton, J., Weissberg, R. P., Durlak, J. A., Dymnicki, A. B., Taylor, R. D., Schellinger, K. B., & Pachan, M. (2008). The positive impact of social and emotional learning for kindergarten to eighth-grade students: Findings from three scientific reviews. Collaborative for Academic, Social, and Emotional Learning. https://files.eric.ed.gov/fulltext/ED505370.pdf

Pulkkinen, L., Nygren, H., & Kokko, K. (2002). Successful development: Childhood antecedents of adaptive psychosocial functioning in adulthood. Journal of Adult Development, 9, 251–265. https://doi.org/10.1023/A:1020234926608

Rose, Evan K., Jonathan T. Schellenberg, and Yotam Shem-Tov. *The Effects of Teacher Quality on Adult Criminal Justice Contact*. No. w30274. National Bureau of Economic Research (2022).

Zimmerman MA, Eisman AB, Reischl TM, et al. Youth Empowerment Solutions: Evaluation of an After-School Program to Engage Middle School Students in Community Change. Health Education & Behavior. 2018;45(1):20-31.