

Evaluation Plan for a District After-School Program

A Quasi-Experimental Mixed-Methods Approach

Everett Stuckey

February 9, 2026

Introduction

This evaluation plan outlines a rigorous approach to assessing the implementation and effectiveness of the district's new after-school program. The plan is divided into two major sections: a formative evaluation to guide program improvement during early implementation, and a summative evaluation to determine the program's overall impact and effectiveness. Within each section, specific topics are addressed in subsections including Purposes, Stakeholders, Evaluation Questions, Design and Methods, Data Collection, Data Analysis, Logistics, and Budget.

Because random assignment of students to participate or not participate in the after-school program is not feasible due to ethical and practical constraints, this evaluation plan employs a quasi-experimental design. Specifically, the summative evaluation will use a nonequivalent comparison group design with propensity score matching (PSM) to approximate the rigor of a randomized controlled trial. This approach allows the research team to draw stronger causal inferences about program effects while respecting the voluntary nature of student participation.

A mixed-methods approach will be used throughout the evaluation. Quantitative methods will be used to measure academic outcomes, attendance, and behavioral indicators, while qualitative methods will capture the experiences and perspectives of students, families, staff, and administrators. Triangulating data from multiple sources and methods will strengthen the validity of the evaluation's conclusions.

Background

The district is planning to implement a new after-school program designed to provide academic support, enrichment activities, and a safe and supervised environment for students outside of regular school hours. The program aims to improve student academic achievement, increase school engagement, reduce risky behaviors, and support working families. Programming may include homework help and tutoring, STEM enrichment, arts and creative activities, physical fitness and recreation, and social-emotional learning (SEL) components.

The after-school program is anticipated to serve students across multiple grade levels and school buildings within the district. Participation is voluntary and open to all students, though targeted outreach will be made to students identified as at-risk based on academic performance, attendance, or behavioral indicators. Because participation is self-selected, the evaluation must account for potential selection bias, students who choose to participate may differ systematically from those who do not. This is the central methodological challenge addressed in this plan.

Research on after-school programs consistently identifies several factors associated with effective programs, including sustained participation (dosage), quality of instruction, alignment with the school-day curriculum, and strong relationships between staff and students (Durlak et al., 2010; Lauer et al., 2006). This evaluation will examine these implementation factors alongside student outcome data to provide a comprehensive picture of program quality and impact.

Evaluation Standards

This evaluation will be guided by the following standards and frameworks:

Program Quality Standards

- Joint Committee on Standards for Educational Evaluation: Program Evaluation Standards (Utility, Feasibility, Propriety, Accuracy, Accountability)
- Collaborative for Academic, Social, and Emotional Learning (CASEL) framework for SEL programming
- Missouri Department of Elementary and Secondary Education (DESE) standards for extended learning programs
- District curriculum standards and student learning outcomes

Research Design Standards

- What Works Clearinghouse (WWC) Standards for quasi-experimental designs
- American Evaluation Association Guiding Principles for Evaluators
- Institute of Education Sciences (IES) guidelines for evidence levels

Formative Evaluation Plan

Purposes

The purpose of the formative evaluation is to monitor and improve the after-school program during its initial implementation. The formative evaluation will focus on program fidelity, quality of implementation, participant engagement, and early indicators of progress. Findings will be used to make real-time adjustments to programming, staffing, scheduling, and content delivery. The primary outcome of the formative evaluation will be a mid-year report with actionable recommendations for program improvement.

Stakeholders

The primary client of the evaluation is the district administration that has requested the research team to prepare this plan. Primary stakeholders include:

- **District Administration:** Decision-makers responsible for program funding, continuation, and expansion.
- **Program Coordinators and Staff:** Individuals directly responsible for delivering after-school programming. Their daily experiences are essential for understanding implementation challenges and successes.
- **Building Principals:** School leaders who oversee logistics, space, and coordination between the school day and after-school activities.
- **Teachers:** School-day teachers whose students participate in the program. They can observe carryover effects in academic performance and behavior.
- **Students:** The direct beneficiaries of the program. Their engagement, satisfaction, and growth are central to the evaluation.
- **Parents/Guardians:** Families who rely on the program and whose support is critical for sustained student participation.

Secondary audiences include the school board, community partners providing supplemental programming, and potential funders or grant agencies that may require evidence of program effectiveness.

Decisions

The results of the formative evaluation will inform decisions about:

Adjustments to program scheduling, staffing ratios, and activity offerings
Targeted interventions for students with low attendance or engagement
Professional development needs for after-school program staff
Communication strategies with families to increase enrollment and retention
Alignment of after-school content with school-day curriculum

Evaluation Questions

The formative evaluation will address the following questions:

1. Is the program being implemented as designed?

- Are planned activities being delivered consistently across sites?
- Are staffing levels and qualifications meeting program standards?
- Is the program schedule being followed as intended?

2. What is the level of student participation and engagement?

- What are daily and weekly attendance rates for the program?
- What is the average dosage (hours per week) for participating students?
- How engaged are students during program activities?

3. What are the experiences and satisfaction levels of key stakeholders?

- How do students perceive the quality and relevance of program activities?
- How do parents rate their satisfaction with the program?
- What challenges and successes do program staff report?

4. What early indicators suggest the program is on track to meet its goals?

- Are participating students showing improvement in homework completion?
- Are there early changes in school-day attendance or behavior for participants?

Participant Sample

Formative data will be collected from all students enrolled in the after-school program. Additionally, a convenience sample of program staff (all staff members) and a stratified random sample of parents/guardians will be surveyed. Program observations will be conducted at each participating school site. The research team will aim for a parent survey response rate of at least 30% to ensure adequate representation.

Methods

The formative evaluation will use a mixed-methods approach including the following data collection strategies:

Data Collection Method	Type	Source	Frequency
Program attendance logs	Quantitative	Program staff	Daily
Structured program observations	Qualitative	Research team	Bi-weekly per site
Student engagement surveys	Quantitative	Students	Monthly
Staff reflection logs and interviews	Qualitative	Program staff	Monthly
Parent satisfaction surveys	Mixed	Parents/Guardians	Mid-year
Fidelity checklists	Quantitative	Research team / Staff	Weekly

Program observations will use a structured protocol assessing the quality of instruction, student engagement, classroom management, and alignment with program goals. The observation tool will be adapted from established instruments such as the Out-of-School Time (OST) observation tool. Staff reflection logs will capture weekly notes on what is working, challenges encountered, and adaptations made.

Logistics

The formative evaluation will begin at program launch and continue through the first semester. The research team will designate one member as the primary liaison with program sites. Observation schedules will be distributed in advance and rotated across sites to minimize disruption. Monthly check-in meetings with program coordinators will be held to share preliminary findings and discuss adjustments. A mid-year formative report will be delivered to district administration by the end of the first semester.

Budget

Item	Estimated Cost
Personnel (research team, ~100 hours @ \$35/hr)	\$3,500
Survey platform (e.g., Qualtrics)	\$500 (or in-kind)
Observation protocol materials and printing	\$200
Incentives for parent survey completion	\$300
Data management software	\$250
Contingency	\$250

Total Estimated Formative Evaluation Budget: \$5,000

Summative Evaluation Plan

Purposes

The purpose of the summative evaluation is to determine the overall effectiveness and impact of the after-school program on student outcomes. The summative evaluation will assess whether the program achieved its intended goals of improving academic achievement, increasing school engagement, improving attendance, and reducing behavioral incidents. The evidence collected will guide decisions about program continuation, expansion, modification, or discontinuation.

Stakeholders

The stakeholders for the summative evaluation are the same as those identified in the formative evaluation plan. The primary audience for summative findings includes district administration and the school board, who will use the results to make resource allocation and program continuation decisions. Summative results will also be shared with program staff, teachers, families, and community partners. If grant funding is involved, summative findings will be reported to funders as evidence of program impact.

Decisions

The results of the summative evaluation will inform the following decisions:

Should the after-school program be continued, expanded, or discontinued?

Is the program producing a statistically significant impact on student academic outcomes?

Which program components are most strongly associated with positive outcomes?

Should additional resources be allocated to scale the program to more schools?

What modifications are needed for the next program cycle?

Evaluation Questions

1. To what extent did the after-school program improve participating students' academic achievement?

- Did participants show greater gains in GPA compared to matched non-participants?
- Did participants show greater gains on standardized assessments (e.g., MAP scores) compared to matched non-participants?
- Is there a dose-response relationship between hours of participation and academic gains?

2. To what extent did the program improve student attendance and reduce chronic absenteeism?

- Did participants have higher school-day attendance rates than matched non-participants?
- Did the program reduce the rate of chronic absenteeism among participants?

3. To what extent did the program improve student behavior and school engagement?

- Did participants have fewer disciplinary referrals than matched non-participants?
- Did participants report higher levels of school connectedness and engagement?

4. What are the experiences and perceptions of students, families, and staff regarding the program's value?

- What do students identify as the most valuable aspects of the program?
- How do parents perceive the program's impact on their child?
- What program features do staff believe contributed most to student success?

5. To what extent was the program implemented with fidelity, and how did implementation quality relate to outcomes?

- Did sites with higher fidelity scores produce better student outcomes?
- What implementation factors were most strongly associated with program success?

Research Design: Quasi-Experimental Approach

Because random assignment of students to treatment (program participation) and control (non-participation) conditions is not feasible, this evaluation will employ a quasi-experimental nonequivalent comparison group design. This is the most rigorous design available when randomization is not possible, and it is recognized by the What Works Clearinghouse (WWC) as capable of producing evidence that "meets standards with reservations" when proper analytic adjustments are made.

Propensity Score Matching (PSM)

To address selection bias, the possibility that students who choose to participate differ systematically from those who do not, the evaluation will use propensity score matching (PSM). Propensity scores represent the probability that a given student would participate in the program based on observed baseline characteristics. Students in the treatment group (program participants) will be matched to students in the comparison group (non-participants) who have similar propensity scores, thereby creating groups that are more comparable on observed covariates.

Covariates used to estimate propensity scores will include:

Baseline GPA (prior year)

Baseline standardized test scores (prior year MAP scores)
Prior-year attendance rate
Number of prior-year disciplinary referrals
Free/reduced lunch eligibility (as a proxy for socioeconomic status)
Grade level
Gender
Race/ethnicity
English Learner (EL) status
Special education (IEP) status
School building

After matching, balance diagnostics will be conducted to verify that the treatment and comparison groups are comparable on all observed covariates. Standardized mean differences of less than 0.25 on each covariate will be the threshold for acceptable balance, consistent with WWC guidelines. If balance is not achieved through one-to-one nearest-neighbor matching, alternative approaches such as caliper matching, stratification on propensity scores, or inverse probability of treatment weighting (IPTW) will be considered.

Supplementary Design Elements

To further strengthen causal inference, the evaluation will incorporate the following supplementary design elements:

Pretest-Posttest Design: Baseline (pretest) measures of all academic and behavioral outcomes will be collected before the program begins. Comparing pre-to-post change between treatment and matched comparison groups strengthens the ability to attribute observed differences to the program rather than preexisting differences.

Dose-Response Analysis: Within the treatment group, the relationship between the amount of program participation (total hours attended) and outcomes will be examined. A dose-response relationship, where more participation is associated with better outcomes, provides additional evidence supporting a causal interpretation.

Difference-in-Differences (DiD): As an additional analytic strategy, a difference-in-differences approach will compare the change in outcomes from pre to post between treatment and comparison groups. This approach controls for any time-invariant unobserved differences between groups.

Sensitivity Analysis: Rosenbaum bounds or similar sensitivity analyses will be conducted to assess how robust the findings are to potential unmeasured confounders. This provides transparency about the limits of causal inference in the absence of random assignment.

Data Collection

A comprehensive set of quantitative and qualitative data will be collected to answer the summative evaluation questions. The following table summarizes the recommended data types, sources, timing, and associated evaluation questions.

Data Type	Specific Measure	Source	Timing	Evaluation Question(s)
Academic Achievement	Cumulative GPA	District SIS	Pre (prior year) & Post (end of year)	1
Academic Achievement	MAP / standardized test scores	State assessment data	Pre & Post	1
Academic Achievement	Course grades (core subjects)	District SIS	Quarterly	1
Attendance	School-day attendance rate	District SIS	Pre & Post (daily)	2
Attendance	Chronic absenteeism indicator	District SIS	Pre & Post	2
Behavior	Disciplinary referrals / suspensions	District SIS	Pre & Post	3
Engagement	Student engagement survey	Students	Pre & Post	3, 4
Program Participation	Daily attendance logs (dosage)	Program staff	Daily	1, 2, 3, 5
Perceptions	Student experience interviews/focus groups	Students (sample)	End of year	4
Perceptions	Parent satisfaction survey	Parents/Guardians	End of year	4
Perceptions	Staff interviews	Program staff	End of year	4, 5
Implementation	Fidelity observation scores	Research team	Ongoing	5

Quantitative Data

Administrative records from the district Student Information System (SIS) will provide the primary quantitative outcome data. These records include GPA, course grades,

standardized test scores, daily attendance, and disciplinary records. These data are collected routinely by the district and require no additional burden on students or staff. Baseline (prior-year) values of these measures will serve as pretests, and current-year values will serve as posttests.

Program participation data, including daily sign-in/sign-out records, activities attended, and total hours of participation, will be collected by program staff using a standardized attendance tracking system. This dosage data is critical for the dose-response analysis.

A validated student engagement survey (e.g., the Student Engagement Instrument or a district-adapted version) will be administered to both program participants and comparison students at the beginning and end of the school year to measure changes in school connectedness, cognitive engagement, and peer/teacher relationships.

Qualitative Data

Semi-structured interviews and focus groups will be conducted with a purposive sample of students, parents, and program staff at the end of the program year. Student focus groups (4-6 groups of 6-8 students each, stratified by grade level and school site) will explore students' experiences, perceived benefits, and suggestions for improvement. Individual interviews with 8-10 parents and 6-8 program staff members will provide additional depth. All interviews and focus groups will be audio-recorded and transcribed for analysis.

Data Analysis Strategies

Quantitative Analysis

Propensity Score Estimation: A logistic regression model will estimate propensity scores using the baseline covariates listed above. Matching will be performed using nearest-neighbor matching (1:1 or 1:k) with a caliper of 0.2 standard deviations of the logit of the propensity score.

Balance Checks: Standardized mean differences and variance ratios will be computed for all covariates before and after matching to verify group equivalence.

Outcome Analysis: For continuous outcomes (GPA, test scores, attendance rates), analysis of covariance (ANCOVA) or multilevel regression models will compare post-program outcomes between matched treatment and comparison groups, controlling for pretest scores and any remaining covariate imbalances. Effect sizes (Cohen's d) will be reported.

Difference-in-Differences: A DiD regression model will estimate the treatment effect as the interaction between group assignment (treatment vs. comparison) and time (pre vs. post), with covariates included.

Dose-Response Analysis: Within the treatment group, regression models will examine the relationship between total hours of participation and outcomes, controlling for baseline characteristics.

Subgroup Analyses: Moderator analyses will examine whether program effects differ by student subgroups (e.g., grade level, baseline achievement level, free/reduced lunch status, race/ethnicity).

Sensitivity Analysis: Rosenbaum bounds will assess sensitivity to hidden bias from unmeasured confounders.

Qualitative Analysis

Interview and focus group transcripts will be analyzed using thematic analysis (Braun & Clarke, 2006). Two members of the research team will independently code transcripts, then meet to reconcile codes and identify emergent themes. Themes will be organized around the evaluation questions. Member checking will be used where feasible to validate interpretations. Open-ended survey responses will be analyzed using the same thematic approach.

Integration of Quantitative and Qualitative Findings

Findings from quantitative and qualitative analyses will be integrated using a convergent parallel mixed-methods design (Creswell & Plano Clark, 2017). Quantitative results will establish the magnitude of program effects, while qualitative findings will provide explanatory context, illuminating how and why the program did or did not produce its intended effects. Joint displays will be used in the final report to present aligned quantitative and qualitative findings side by side.

Matrix of Summative Questions and Data Collection Methods

Evaluation Question	Data Source(s)	Method	Analysis
1. Academic achievement impact	GPA, MAP scores, course grades	PSM + ANCOVA / DiD	Regression, effect sizes
2. Attendance and chronic absenteeism	SIS attendance records	PSM + ANCOVA / DiD	Regression, chi-square
3. Behavior and engagement	Disciplinary records, engagement survey	PSM + ANCOVA; Pre/post survey	Regression, paired t-tests
4. Stakeholder perceptions	Interviews, focus groups, surveys	Qualitative; Descriptive stats	Thematic analysis, frequencies
5. Implementation fidelity and outcomes link	Fidelity scores, dosage, outcomes	Correlation / regression	Multilevel modeling

Participant Sample

The treatment group will consist of all students who enroll in and attend the after-school program for a minimum threshold of participation (e.g., at least 30 days or 60 hours over the school year) to ensure a meaningful dosage. The comparison group will be drawn from students in the same schools and grade levels who did not participate in the program. Propensity score matching will be used to select comparison students who are most similar to participants on baseline characteristics.

For qualitative data collection, purposive sampling will be used to select students and parents who represent a range of experiences (e.g., high-dosage vs. low-dosage participants, students from different grade levels and school sites). Program staff from each site will be interviewed.

A power analysis should be conducted once enrollment numbers are known to determine whether the sample size is sufficient to detect a minimum effect size of practical significance (e.g., Cohen's $d = 0.20$) with 80% power at $\alpha = 0.05$. Based on anticipated enrollment of 200-300 participants, a 1:1 matched design should provide adequate power for the primary outcome analyses.

Logistics

Summative data collection will span the full program year. Baseline administrative data (prior-year GPA, attendance, test scores, disciplinary records) will be extracted from the district SIS at the start of the school year. The student engagement survey will be administered in September (pretest) and May (posttest) to both treatment and comparison groups. Program attendance logs will be collected continuously throughout the year. End-of-year interviews and focus groups will be conducted in May. Administrative outcome data will be extracted after the close of the school year in June.

Team Responsibilities

Task	Responsible Party	Timeline
Baseline data extraction and PSM	Lead Researcher	September-October
Survey administration (pre/post)	Research Assistant	September & May
Program observation and fidelity monitoring	Research Team	Ongoing
Interviews, focus groups, and qualitative analysis	Qualitative Lead	May-June

A final summative evaluation report will be delivered to district administration by August, allowing findings to inform decisions about the next program year. The report

will include an executive summary, detailed findings organized by evaluation question, limitations, and recommendations.

Budget

Item	Estimated Cost
Personnel (research team, ~250 hours @ \$35/hr)	\$8,750
Survey platform (Qualtrics or equivalent)	\$500 (or in-kind)
Statistical software (e.g., R/Stata, may be in-kind)	\$500
Transcription services for interviews/focus groups	\$1,200
Incentives (student focus groups, parent interviews)	\$800
Printing and materials	\$250
Data management and security	\$500
Contingency	\$500

Total Estimated Summative Evaluation Budget: \$13,000

Total Combined Evaluation Budget (Formative + Summative): \$18,000

Note: Costs may be reduced if the district can provide in-kind contributions such as existing survey software licenses, statistical software, and personnel time allocated from existing research office budgets.

Limitations and Threats to Validity

While the quasi-experimental design with propensity score matching represents the most rigorous approach available without random assignment, several limitations must be acknowledged:

Selection Bias from Unobserved Confounders: PSM can only account for observed covariates. Unmeasured factors, such as student motivation, parental involvement, or peer influence, may still differ between groups. Sensitivity analyses (Rosenbaum bounds) will quantify how large such hidden bias would need to be to overturn findings.

Attrition: Students may drop out of the program during the year, potentially introducing attrition bias. Intent-to-treat (ITT) analysis will be used as the primary analytic approach, supplemented by treatment-on-the-treated (TOT) analysis for students meeting the dosage threshold.

Generalizability: Findings are specific to this district's context and may not generalize to other settings without replication.

Contamination: Comparison group students may participate in other after-school activities or tutoring, which could dilute the estimated program effect. Data on comparison students' out-of-school activities will be collected where feasible.

Implementation Variability: Differences in implementation quality across school sites may affect outcomes. Multilevel models that account for site-level clustering will be used to address this.

Appendix A: Recommended Data Collection Instruments

Student Engagement Survey (Pre/Post)

Administered to both treatment and comparison groups at the beginning and end of the school year. Items rated on a 5-point Likert scale (1 = Strongly Disagree to 5 = Strongly Agree).

1. I feel like I belong at my school.
2. I pay attention in my classes.
3. I try hard to do well in school.
4. I feel safe at school.
5. I have at least one adult at school I can talk to if I have a problem.
6. I am interested in what I am learning at school.
7. I feel motivated to come to school each day.
8. I believe school is preparing me for my future.
9. I participate actively in school activities.
10. I complete my homework on time.

Parent/Guardian Satisfaction Survey

Administered at the end of the school year to parents of program participants.

1. How satisfied are you with the after-school program overall?

Response options: Very Satisfied / Satisfied / Neutral / Dissatisfied / Very Dissatisfied

2. My child enjoys attending the after-school program.

Response options: Strongly Agree / Agree / Neutral / Disagree / Strongly Disagree

3. The program has helped my child academically.

Response options: Strongly Agree / Agree / Neutral / Disagree / Strongly Disagree

4. The program provides a safe environment for my child.

Response options: Strongly Agree / Agree / Neutral / Disagree / Strongly Disagree

5. I would recommend this program to other families.

Response options: Strongly Agree / Agree / Neutral / Disagree / Strongly Disagree

6. What do you see as the biggest benefit of the program for your child?

Response options: [Open-ended response]

7. What improvements would you suggest for the program?

Response options: [Open-ended response]

8. Has the program made it easier for you to manage work or other responsibilities?

Response options: Yes / Somewhat / No

Student Focus Group Protocol

Semi-structured focus group guide for end-of-year student focus groups (45-60 minutes, 6-8 students per group).

1. Tell me about your experience in the after-school program this year. What was a typical day like?
2. What activities or parts of the program did you enjoy the most? Why?
3. Were there any activities or parts of the program you did not enjoy? Why?
4. Do you think the program helped you with your schoolwork? Can you give an example?
5. Did you feel safe and supported in the after-school program? Tell me more.
6. How would you describe the adults who work in the after-school program?
7. Did your feelings about school change because of the program? How?
8. If you could change one thing about the program, what would it be?
9. Would you recommend the program to a friend? Why or why not?
10. Is there anything else you would like to share about your experience?

Program Staff Interview Protocol

Individual semi-structured interviews with program staff (30-45 minutes each).

1. Describe your role in the after-school program and your typical responsibilities.
2. What do you see as the primary strengths of the program?
3. What challenges have you encountered in implementing the program?

4. How would you describe student engagement and participation over the course of the year?
5. Have you observed any changes in students' academic behaviors (e.g., homework completion, attitude toward school)?
6. How well do you feel the after-school program aligns with what students are learning during the school day?
7. What professional development or support would help you be more effective in your role?
8. What recommendations would you make for improving the program next year?

Program Fidelity Observation Checklist

Completed by the research team during structured program observations. Each item rated on a 4-point scale (1 = Not Observed, 2 = Partially Observed, 3 = Mostly Observed, 4 = Fully Observed).

1. Activities align with the program's stated goals and curriculum.
2. Staff-to-student ratio meets program standards.
3. Program begins and ends at the scheduled times.
4. Students are actively engaged in program activities.
5. Staff use positive and supportive interactions with students.
6. Academic support activities (e.g., homework help, tutoring) are available.
7. Enrichment activities are offered as planned.
8. A safe and organized environment is maintained.
9. Staff redirect off-task behavior effectively and respectfully.
10. Materials and resources are available and adequate.

Appendix B: Summary of Recommended Data to Collect

The following is a consolidated summary of all data types recommended for collection in this evaluation:

Administrative / Existing Data (from District SIS)

Student demographics (grade, gender, race/ethnicity, FRL status, EL status, IEP status)
Prior-year and current-year GPA
Prior-year and current-year standardized test scores (e.g., MAP)
Prior-year and current-year daily attendance records
Prior-year and current-year disciplinary referrals and suspensions
Course grades by subject (quarterly)

Program-Generated Data

Daily program attendance (sign-in/sign-out logs)
Activities attended by each student
Total hours of participation per student (dosage)
Program fidelity observation scores
Staff reflection logs

Survey Data

Student engagement survey (pre and post, treatment and comparison groups)
Parent/guardian satisfaction survey (post, treatment group)
Student monthly engagement check-ins (formative)
Staff surveys (formative)

Qualitative Data

Student focus group transcripts
Parent interview transcripts
Program staff interview transcripts
Open-ended survey responses
Structured observation field notes

References

- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77-101.
- Creswell, J. W., & Plano Clark, V. L. (2017). Designing and conducting mixed methods research (3rd ed.). SAGE.
- Durlak, J. A., Weissberg, R. P., & Pachan, M. (2010). A meta-analysis of after-school programs that seek to promote personal and social skills in children and adolescents. *American Journal of Community Psychology*, 45(3-4), 294-309.
- Joint Committee on Standards for Educational Evaluation. (2011). The program evaluation standards (3rd ed.). SAGE.
- Lauer, P. A., Akiba, M., Wilkerson, S. B., Apthorp, H. S., Snow, D., & Martin-Glenn, M. L. (2006). Out-of-school-time programs: A meta-analysis of effects for at-risk students. *Review of Educational Research*, 76(2), 275-313.
- Rosenbaum, P. R. (2002). Observational studies (2nd ed.). Springer.
- Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1), 41-55.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). Experimental and quasi-experimental designs for generalized causal inference. Houghton Mifflin.
- U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse. (2020). What Works Clearinghouse standards handbook (Version 4.1).