

An Independent Evaluation of Global Game Changers’ Impact on Student Outcomes in Muhlenberg County Public Schools during the 2024-2025 School Year

Abstract

This report presents an independent evaluation of a quasi-experimental study of Global Game Changers Children’s Education Initiative’s (GGC) curriculum implemented during the 2024–2025 school year in Muhlenberg County Public Schools. Outcomes were compared between one elementary school implementing the program and a demographically similar control school in the same district that maintained normal instruction. Student self-report surveys (grades 3–5) and teacher observation surveys (K–5) measured outcomes on leadership, emotional, and social skills. After adjusting for baseline differences and grade level, students in the treatment school reported significantly higher post-intervention leadership, emotional, and social skills score ($\beta = 7.87$, $SE = 2.33$, $p < .001$). Teacher observation scores also favored the treatment school ($\beta = 1.94$, $SE = 1.13$, $p = .085$). Overall, findings are consistent with expected modest, positive first-year effects associated with program implementation.

Prepared for: Global Game Changers Children’s Education Initiative, Inc.

Independent Evaluator: Jarett S. Lopez

Introduction

Student behaviors, attitudes, and willingness to engage in peer leadership are of key interest to educators and service providers. A growing body of research links programming on leadership, social, and emotional skills to improvements in student interpersonal relationships, academic engagement, and school climate (Durlak et al., 2011; Taylor et al., 2017). As a result, schools are increasingly asked not only to implement programs that enrich these leadership, emotional, and social skills, but also to demonstrate that those programs produce measurable and meaningful outcomes.

Global Game Changers Children’s Education Initiative (henceforth GGC) is a leadership, emotional, and social skills development nonprofit whose program is designed for students in pre-kindergarten through fifth grade. The curriculum emphasizes whole-child development through a structured programming that helps students identify their strengths, connect those strengths to leadership, emotional, and social skills, and apply them in classroom and community contexts. This approach aligns with broader social emotional behavior frameworks that emphasize skills development, self-awareness, and social responsibility (Collaborative for Academic, Social, and Emotional Learning [CASEL], 2020).

This report presents an independently evaluated quasi-experiment of GGC’s program during the 2024–2025 academic year in Muhlenberg County Public Schools, a rural district in western Kentucky. Two elementary schools participated in the study: Muhlenberg South Elementary, which implemented GGC curriculum, and Bremen Elementary, which served as a comparison school. Evaluations of programs that support leadership, emotional, and social skills in rural contexts remain comparatively limited, despite

evidence that rural districts often face distinct capacity and implementation challenges (Showalter et al., 2017).

The evaluation draws on complementary but distinct outcome measures. Student self-report surveys capture changes in perceived leadership, emotional, and social skills among students in grades three through five, while teacher observation surveys provide an external behavioral perspective across elementary grade levels. Using multiple informants allows for triangulation of findings and reduces reliance on any single measurement source (De Los Reyes & Kazdin, 2005). After accounting for baseline differences where applicable, students reported higher post-intervention leadership, emotional, and social skill scores. Teacher observation scores suggested a similar relationship but fell short of conventional levels of statistical significance.

2. Research Design

2.1 Quasi-Experimental Studies

A quasi-experimental design estimates treatment effects under non-random assignment by comparing outcomes between treatment and control groups while adjusting for baseline differences. The QED constructed in this case includes two elementary schools in Muhlenberg County, Kentucky. Muhlenberg county is a rural county in southwest Kentucky with approximately 31,000 residents according to 2020 census data.

2.2 Group Assignment

Muhlenberg South Elementary School was the treatment case implementing GGC curriculum intervention in its kindergarten through 5th grade classrooms. Bremen, the control school received no exposure to any component of GGC curriculum, meeting CASEL's operational definition of a control group. These assignments occurred through GGC collaboration with district leadership to participating schools. Student data across measures is matched at the pre and post levels where possible.

2.3 Intervention/Treatment Curriculum

The curriculum serving as the treatment in this quasi-experimental design is a PreK-5th grade leadership, emotional, and social skills program. It provides fully developed lessons aligned with CASEL competencies that emphasize whole-child development, student voice, and leadership skill development through real-world applicable project-based activities.

The curriculum is centered on the Global Game Changers Superpower Equation – MY TALENT + MY HEART + MY SUPERPOWER!®, which encourages students to explore what they care about (Heart), uncover their unique strengths (Talent), and build the self-confidence they need to thrive in school

and in life. Every lesson helps kids connect their talents to causes they love, nurturing empathy and a sense of purpose through discovering their unique Superpowers. These elements collectively represent the core GGC treatment condition examined in this evaluation. Students at the comparison school received no exposure to any component of the GGC curriculum. Instead, Bremen maintained typical instruction.

2.4 Statistical Tests and Analytic Sequencing

The analytic plan is structured to align with CASEL and What Works Clearinghouse standards for evaluating student interventions using quasi-experimental designs. Baseline equivalence is assessed descriptively using t-tests and standard mean differences, while ANCOVA models serve as the primary method for estimating adjusted treatment effects. Linear models were used for student survey and teacher observation scores, and a logistic model was used for the binary improvement indicator. This multivariate approach is intended to ensure a broad base of data with which to determine treatment impact on participating students and their leadership, emotional, and social skills.

Analyses draw on school and district indicators, student pre/post surveys, and Teacher Observation Surveys to maintain consistency across data sources. Data derived from a post-implementation instructor focus group is also included, providing additional context on the student surveys and teacher observation surveys. T-tests are used descriptively within the student surveys and to suggest direction/sign of relationship between the groups before controls for baseline differences are applied. Although multilevel modeling was considered, the small number of schools ($n = 2$) precluded reliable estimation of school-level random effects; accordingly, adjusted single-level models were used.

2.5 Variables

GGC and its partners at the school and district level provided a robust set of data with which to evaluate program impact. These data support triangulation across multiple sources and include student, teacher, and school-derived measures.

2.5.1 School and District Level Variables

The Kentucky Department of Education requires Muhlenberg County Public Schools to collect and report a variety of data to the Commonwealth. Relevant data, including enrollment, Title I eligibility, student demographic composition, and relevant school metrics, including student-to-teacher ratio, instructor turnover, and per-pupil spending are included.

Figure 1 below gives a descriptive presentation of the similarities of Muhlenberg South and Bremen Elementaries at baseline as schools. This is relevant as a differently resourced school or one with a composition significantly different from the other would bias results (King, Keohane, and Verba, 1994).

Fig. 1: Muhlenberg South and Bremen Elementary, Descriptive Statistics, 2024-25 Academic Year		
Variable	Muhlenberg South	Bremen
Enrollment	489 students	323 students
Title I Eligible	Yes	Yes
% Economic Disadvantaged	70.1%	62.2%
% White	94.7%	92.6%
Minority Plurality	Two or More Races	Two or More Races
Student to Teacher Ratio	13:1	12:1
Teacher Turnover %	14.7%	3.7%
Per Pupil Spending ALL (23-24)	\$13,440	\$15,817
Source: Kentucky Department of Education, Report Card Dashboards 2023-2024 and 2024 -2025 Academic Year. https://reportcard.kyschools.us/		

In this case, at the outset of the school year, Muhlenberg South and Bremen Elementaries have negligible institutional differences. Marginal differences that do exist can be explained by a slightly smaller enrollment, do not have an impact the analysis, and cannot be distinguished from treatment effects at $n = 2$ (Shadish, Cook, and Campbell, 2002).

2.5.2 Pre/Post GGC Student Surveys

In the treatment and in the control school, students in 3rd - 5th grades completed a survey responding to a series of 18 statements designed to gauge key internal and behavioral leadership, emotional, and social skills (Appendix I). Each statement is rated on a scale of Never (0) to Always (4). Students completed the pre-survey in October 2024; post-surveys were completed in April 2025. Implementing instructors administered the survey which were collected, scanned, and delivered to the independent evaluator for cleaning, coding, and processing in R Statistical Computing.

The answers to these questions are then summed into a leadership, emotional, and social skill score 0 to 72. Pre - and post - survey results are examined and reported within and between groups at baseline and outcome. This allows insight and statistical inference on student perception of their own leadership, emotional, and social skills. (West, Pier, Fricke, et al, 2020).

2.5.3 Teacher Observation Surveys

Teacher Observation Surveys (Appendix II) provide a structured way for teachers to rate observable student behaviors related to leadership, emotional, and social skills, including positive emotional expression, the ability to articulate needs, cooperation with peers, and general classroom participation. These data serve as an external behavioral indicator to complement the student perspective gathered from the surveys. In rating each student by name, teachers respond to each item on a four-point frequency scale from Never to Always, producing a measure of student behavior suitable for descriptive and comparative analysis.

2.5.4 Student Characteristics

Individual-level student characteristics are included as the first base-level comparison before examining survey data. These measures include demographic and socioeconomic control measures like gender, IEP eligibility, Free/Reduced Price Lunch eligibility, student age, and absenteeism proportion indicate that students are equivalent on indicators that have statistically significant relationship with access to learning opportunities and downstream academic and behavioral outcomes. These variables are included due to their propensity to shape opportunity structure and bias simple posttest results (Stuart, 2010).

2.5.5 Additional Qualitative Data

Interpretive data is derived from an end-of-implementation focus group with treatment group instructors and administrative staff. The data derived from that focus group is an additional way to add context and nuance to the formal statistical analysis of the intervention. A thematic analysis of this conversation is included as a method to gauge treatment implementation fidelity and to provide anecdotal data on student behavioral impacts from the beginning to the end of the academic year.

2.6 Attrition and Missingness

Not all measures were available for all students. For example, student self-report surveys are administered only to students in grades 3rd–5th. This is the best practice, as young children have difficulty distinguishing ordinal response categories on Likert-type scales (Carifio and Perla, 2008). Instead, Teacher Observation Surveys are utilized for these students and provide additional outcome data on surveyed students. Measurement fatigue for both the respondents and the implementers is also a risk, so the selective data collection approach minimizes this possibility (King, Keohane, and Verba, 1994)

Analytic attrition in this study is operationalized as students who were not surveyed at either temporal point. Students without post-survey data cannot be included in outcome analyses because there is

no comparable follow-up measure from which to infer changes in their self-reported leadership, emotional, and social skills. Likewise, students with only a post-survey are excluded because they joined the treatment environment after the baseline condition was established; therefore, the extent of their exposure to GGC curriculum cannot be determined, and no valid change estimate can be produced. A statistical measure of missingness is included in the results section. True attrition from the treatment or control groups would be a student moving to a school not included in the analytic sample. The remaining analytic sample, comprising the treatment and control groups, across all measures is also sufficiently large ($n > 500$) to support statistical inference (King, Keohane, and Verba 1994).

3. At Pre-Intervention: Fall 2024

3.1 Student Characteristics

Anonymized, student-level data was obtained from the Muhlenberg County School District. Several notable coding decisions provide context. Demographic indicators for gender, race, and IEP-eligibility were coded as binary variables. For gender, it is framed as if the participant is female, for race the measure represents a non-white student, and IEP-eligibility represents that status. Free/Reduced Price lunch eligibility is also coded on a binary as well as a simple measure of socioeconomic status. Age is treated as a discrete measure in the study with a simple average of ages performed. Finally, absenteeism is a percentage converted back to a proportion of missed days / enrolled days.

For each measure, standardized mean differences (SMDs) are calculated for each of these control variables for both schools. SMDs below 0.25 indicate evidence of substantive baseline equivalence.

See Figure 2 below.

Fig. 2: Demographic and Socioeconomic Treatment and Control Groups			
Variable	Treatment Mean	Control Mean	SMD
Female	0.47	0.43	0.09
Non-White	0.05	0.07	-0.08
IEP-Eligible	0.36	0.41	-0.09
FR/PL	0.78	0.73	0.13

Age	8.42	8.46	−0.02
Absenteeism	0.03	0.02	0.13
<i>Note: Standardized mean differences (SMDs) calculated using pooled standard deviations. Values < 0.25 meet WWC criteria for baseline equivalence. Non-Public data obtained from Muhlenberg County Schools, effective 2024-2025 School Year.</i>			

All standardized mean differences are well below the accepted threshold that indicates that in terms of the included controls. Baseline equivalence standards appear to be met on demographic and socioeconomic indicators.

3.2 Student Surveys

At the start of the study, students completed a survey measuring their leadership, emotional, and social skills (see Appendix I). These baseline results provide important context for interpreting post-intervention outcomes.

Initial comparisons showed that students in the control school reported higher levels of leadership, emotional, and social skills before the program began. This difference was statistically significant, $t(360.16) = -4.24, p < .001$. Considering mean differences, students in the treatment group scored 43.57, compared to 50.19 in the control group—a difference of 6.62 points (95% CI [−9.69, −3.55]). This indicates that the two groups were not equivalent on the student-reported outcome at baseline.

The size of this difference was moderate in magnitude (Cohen's $d = -0.40$), suggesting a meaningful pre-intervention gap rather than a trivial one. For this reason, baseline survey scores were incorporated into all outcome analyses. Adjusted models, including the ANCOVA and gain-score analyses, account for these initial differences so that post-intervention comparisons reflect program-related change rather than pre-existing disparities between schools.

3.3 Teacher Observation Scores

Implementing and non-implementing instructors completed observation surveys for their students based on behavior over the month preceding scoring. Pre-intervention teacher observation scores were nearly identical between students in the treatment and control schools. The difference was not statistically significant ($t(620.35) = 0.67, p = .506$), and the standardized mean difference was very small (SMD = −0.05). In practical terms, students entered the study at comparable levels on the teacher-observed behavior measure. Baseline score is included in later models for comparability with the student survey ANCOVA approach.

3.4 Baseline Equivalence

These findings indicate baseline equivalence on demographic characteristics and teacher-reported behaviors, with a meaningful baseline difference in student-reported leadership, emotional, and social skills scores that is accounted for in subsequent adjusted analyses. These differences are controlled for in the subsequent results section.

4. Implementation Focus Group

An end-of-implementation focus group was conducted with instructors and support staff who participated in delivering the Global Game Changers (GGC) curriculum during the 2024–2025 academic year. The focus group was conducted in May, after all post-surveys were administered and collected.

The purpose of the focus group was to gather descriptive information about how the program was implemented and how instructors experienced its use in practice. These data are included to provide context for the quantitative analyses and are not used to estimate program impact.

The focus group was held via Zoom in May 2025. Analysis consisted of a close reading of the transcript to identify recurring topics raised across participants. Responses were summarized descriptively, with attention to points of agreement rather than isolated comments.

Participants described extensive professional experience in education. Several instructors referenced long teaching tenures and prior exposure to multiple instructional initiatives. This experience shaped how participants discussed GGC, often framing it in comparison to other programs rather than as a new or unfamiliar intervention.

Among the instructors, there was a consistent emphasis on the curriculum as straightforward to implement. They noted that lessons were clearly structured and could be incorporated into existing classroom routines without extensive preparation. GGC's program and intervention were usable to instructors within typical classroom constraints, which supported continued use across the school year.

When discussing implementation, participants frequently referenced student engagement. Instructors reported that students were receptive to the curriculum and willing to participate in discussions prompted by the lessons. Several participants noted that students appeared more comfortable talking about emotions, interpersonal issues, and personal strengths over time. These observations were shared informally and were not quantified, but they reflect instructor's perceptions of how students interacted with the program content. One instructor noted a student-student conflict that was resolved with emotional grounding and other regulation strategies.

The primary issue flagged among participants was time constraint. For instructors, fitting lessons into an already full school day could be difficult. However, they generally described adjusting pacing or timing rather than discontinuing use of the curriculum. No participants described abandoning implementation or being unable to deliver the program as planned. In fact, instructors recommended expanded digital tools to aid in implementation for future GGC-implementation.

Overall, the focus group suggests that GGC curriculum was implemented consistently and was generally supported by instructors. Implementing instructors described the program as feasible within their classrooms and compatible with their teaching practices. These findings indicate that instructors were able to deliver the curriculum as intended within typical classroom constraints, reducing concerns that observed student outcomes were driven by substantial implementation failure rather than program content.

5. Post Intervention: Spring 2025

5.1 Student Survey Data

Initial comparison of the treatment and control groups' outcome data begins with analysis of the student survey data.

5.2 Within Group Analyses

We first examined changes over time within each group to understand overall trends before conducting more rigorous group analyses. Among students in treatment schools, average scores increased slightly from pretest to posttest; however, this change was not statistically significant, $t(243) = 0.94$, $p = .347$. The estimated mean increase was modest ($M_{diff} = 1.43$; 95% CI $[-1.56, 4.42]$), suggesting little measurable change over the study period.

In contrast, students in control schools experienced a statistically significant decline from pretest to posttest, $t(122) = -4.34$, $p < .001$. On average, scores decreased by more than nine points ($M_{diff} = -9.12$; 95% CI $[-13.28, -4.96]$). These within-group patterns provide an initial picture of how outcomes changed over time.

Inferences cannot be made about program impact because of the difference at baseline this t-test indicates that there are group differences that first must be accounted for. For this reason, we rely on adjusted between-group analyses (ANCOVA) in the following section to more accurately estimate program effects.

5.3 ANCOVA and Between Group Analyses

Post-intervention behavioral outcomes were examined using an analysis of covariance (ANCOVA), which adjusts posttest scores for students' baseline leadership, emotional, and social skills as well as grade. This approach allows for a fair comparison between treatment and control students while accounting for pre-existing differences.

After adjustment, participation in the program was associated with significantly higher leadership emotional, and social skills scores. Treatment status was a statistically significant predictor of posttest outcomes ($\beta = 7.87$, $SE = 2.33$, $p < .001$). On average, students in the treatment school scored approximately 8 points higher at post-intervention than students in the control school.

Baseline leadership, emotional, and social skills scores were also a strong predictor of post-intervention outcomes ($\beta = 0.29$, $SE = 0.06$, $p < .001$), indicating that students' initial score continued to influence outcomes over time. Grade-level patterns varied: fourth-grade students did not differ significantly from the reference group, while fifth-grade students scored significantly lower at posttest ($\beta = -8.90$, $SE = 2.75$, $p = .001$).

Overall, the model was statistically significant ($F(4, 362) = 11.41$, $p < .001$) and explained approximately 10% of the variation in post-intervention leadership, emotional, and social skills scores. Full model results are provided in Appendix III.

5.4 Logistic Regression

As an additional layer of analysis, a logistic regression model was estimated to examine the likelihood of improvement at post-intervention as a function of treatment status, baseline score, and grade level. This model tests the null hypothesis that, after adjustment, treatment status is not associated with the odds of improvement.

After adjustment, treatment status was not a statistically significant predictor of improvement ($OR = 1.42$, $p = .150$). Accordingly, the null hypothesis of no adjusted treatment effect on the probability of improvement could not be rejected.

Baseline score was a strong predictor of improvement ($OR = 0.96$, $p < .001$), indicating that students with higher baseline scores were less likely to meet the improvement threshold. Grade-level effects were modest: fourth grade students showed marginally higher odds of improvement relative to the reference group ($OR = 1.67$, $p = .051$), while fifth grade students did not differ significantly ($OR = 0.73$, $p = .292$).

5.5 Teacher Observation Surveys

Post-intervention outcomes were analyzed using an ANCOVA, with post-intervention teacher observation scores regressed on treatment status, baseline score, and grade. This approach allows adjusted

post-intervention differences to be estimated while accounting for initial student behavioral conditions and structural differences across grade levels.

After adjustment, treatment status was associated with higher post-intervention teacher observation scores; however, this difference did not reach conventional levels of statistical significance ($\beta = 1.94$, $SE = 1.13$, $p = .085$, HC3). On average, students in the treatment school scored approximately two points higher than students in the control school, holding baseline scores and grade constant. Baseline observation score remained a strong predictor of post-intervention outcomes ($\beta = 0.70$, $p < .001$), reflecting substantial stability in teacher-observed behavior over time. Students that began the year with higher observation scores tended to remain higher at posttest. Grade level was also a statistically significant predictor ($\beta = 0.98$, $p = .031$), suggesting modest variation in teacher-reported behavior across grades. Including grades in the model helps ensure that grade-related differences in student interpersonal dynamics are not misattributed to treatment effects. The ANCOVA explained approximately 9% of the variance in post-intervention teacher observation scores (adjusted $R^2 = .09$).

To assess whether these findings depended on model specification, a gain-score model was also estimated. Results were directionally consistent with the ANCOVA, with treatment students exhibiting larger average pre-to-post gains than control students; however, the treatment effect again did not reach statistical significance ($\beta = 2.01$, $SE = 1.14$, $p = .077$). Grade was not a statistically significant predictor of gains in this specification. Teacher observation results indicate a modest, directionally positive association between program implementation and observed classroom behavior.

When viewed alongside student-reported social-emotional outcomes, these findings provide a clear positive correlation between program participation when compared to no program participation when adjusting. Student survey results show clearer adjusted differences favoring the treatment group, while teacher-reported behaviors reflect more modest and less differentiated patterns of change. The following discussion situates these findings within the program's theory of change and the broader literature on interventions that improve student well-being as well as leadership, emotional, and social skills.

6. Discussion and Conclusion

This evaluation examined whether participation in Global Game Changers' program was associated with differences in student leadership, emotional, and social skills including teacher-reported classroom behaviors. Students experiencing the GGC curriculum had a clear, positive, and statistically significant outlook based on Pre/Post GGC Survey results. A 7-point increase in student leadership, emotional, and social skills score represents a student who has a stronger sense of global impact, is able to regulate emotionally more effectively, has an optimistic outlook on their future, and is overall happy to be in school. Teacher observation results are a more subtle signal but not contradictory. While differences between

schools' Teacher Observation Survey score consistently favored the treatment school across models but failed to reach statistical significance. Though, that is not to say that the lack of significance implies lack of substance, as a p-value of .085 is not an extreme result that would suggest random chance. A larger future analytic sample including an additional treatment or control school could smooth the noise in the signal.

The alignment in direction across student and teacher outcomes strengthens confidence in the overall pattern of findings, despite the p-value being greater than .05. These findings also raise the question about cumulative impact of Global Game Changers' curriculum. It is designed for grades pre-K-5th grade, with the curriculum building upon itself parallel to the students as they progress through primary and elementary education. A single year implementation across 5 grades is illustrative of initial, positive impact from no-GGC to a year of GGC, but not what the long-term impact is of GGC participation. As GGC has been operating in Kentucky, neighboring states, and across the country for over a decade, there is an opportunity to do this longitudinal impact work, though that level of research may be a challenging endeavor (and beyond the scope of the current project).

The results show that Global Game Changers appears to have a positive impact on student leadership, emotional, and social skills. Though, rather than producing dramatic short-term changes, which would be unrealistic, the program appears to support incremental improvements in these school and program-desired behaviors. The positive pattern present aligns with the broader literature in this area of student leadership, emotional, and social skill development, which emphasizes cumulative and context-dependent effects rather than immediate transformation (Durlak et al., 2011). This evaluation indicates that participation in the Global Game Changers program is associated with modest but consistent improvements in leadership, emotional, and social skills outcomes in a rural elementary school context.

After adjusting for baseline differences and grade-level variation, treatment students reported higher levels of leadership, emotional, and social skills at post-intervention, and treatment classrooms demonstrated non-significant but slightly stronger teacher-reported behavioral outcomes. The pattern is consistent with expectations for education intervention implementation within a single school year. These findings support the conclusion that GGC's curriculum contributes positively to student leadership, emotional, and social skills relative to typical instruction. Continued implementation and longitudinal follow-up would help clarify whether observed behavioral trends strengthen over time.

Bibliography

- Allison, Paul D. "Change Scores as Dependent Variables in Regression Analysis." *Sociological Methodology* 20 (1990): 93–114.
- Carifio, James, and Rocco J. Perla. "Resolving the 50-Year Debate around Using and Misusing Likert Scales." *Medical Education* 42, no. 12 (2008): 1150–52.
- Collaborative for Academic, Social, and Emotional Learning (CASEL). *CASEL Guide to Schoolwide Social and Emotional Learning*. Chicago: CASEL, 2020.
- De Los Reyes, Andres, and Alan E. Kazdin. "Informant Discrepancies in the Assessment of Childhood Psychopathology: A Critical Review, Theoretical Framework, and Recommendations for Further Study." *Psychological Bulletin* 131, no. 4 (2005): 483–509.
- Durlak, Joseph A., Roger P. Weissberg, Allison B. Dymnicki, Rebecca D. Taylor, and Kriston B. Schellinger. "The Impact of Enhancing Students' Social and Emotional Learning: A Meta-Analysis of School-Based Universal Interventions." *Child Development* 82, no. 1 (2011): 405–32.
- King, Gary, Robert O. Keohane, and Sidney Verba. *Designing Social Inquiry: Scientific Inference in Qualitative Research*. Princeton, NJ: Princeton University Press, 1994.
- MacCallum, Robert C., Shaobo Zhang, Kristopher J. Preacher, and Derek D. Rucker. "On the Practice of Dichotomization of Quantitative Variables." *Psychological Methods* 7, no. 1 (2002): 19–40.
- Shadish, William R., Thomas D. Cook, and Donald T. Campbell. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston: Houghton Mifflin, 2002.
- Showalter, David, Richard Klein, Jessica Johnson, and Samantha Hartman. *Why Rural Matters 2015–2016: Understanding the Changing Landscape*. Washington, DC: Rural School and Community Trust, 2017.

- Stuart, Elizabeth A. “Matching Methods for Causal Inference: A Review and a Look Forward.” *Statistical Science* 25, no. 1 (2010): 1–21.
- Taylor, Rebecca D., Eva Oberle, Joseph A. Durlak, and Roger P. Weissberg. “Promoting Positive Youth Development through School-Based Social and Emotional Learning Interventions: A Meta-Analysis of Follow-Up Effects.” *Child Development* 88, no. 4 (2017): 1156–71.
- U.S. Department of Education, Institute of Education Sciences. *What Works Clearinghouse Standards Handbook*, Version 4.1. Washington, DC: U.S. Department of Education, 2020.
- West, Martin R., Laura Pier, Heather Fricke, Heather Hough, Susanna Loeb, Robert H. Meyer, and Andrew B. Rice. “Trends in Student Social-Emotional Learning: Evidence from the First Large-Scale Panel Student Survey.” *Educational Evaluation and Policy Analysis* 42, no. 2 (2020): 279–303

Appendix I. GGC Pre/Post Student Survey ©



Student Name: _____ Grade: _____

Teacher: _____ School: _____

#	Question	Never 0	A little 1	Some- Times 2	A lot 3	Always 4
1	I help other kids.	0	1	2	3	4
2	I like the people at my school.	0	1	2	3	4
3	I have a talent.	0	1	2	3	4
4	I can talk about my feelings.	0	1	2	3	4
5	I can solve problems.	0	1	2	3	4
6	I think the world is a good place.	0	1	2	3	4
7	I can make the world a better place.	0	1	2	3	4
8	I ask for help when I need it.	0	1	2	3	4
9	I can get through hard times.	0	1	2	3	4
10	How I act changes how other people act or feel.	0	1	2	3	4
11	I am happy at school.	0	1	2	3	4
12	I can be a leader.	0	1	2	3	4
13	I understand other people's feelings.	0	1	2	3	4
14	I can make myself feel better when I feel mad/sad.	0	1	2	3	4
15	It's okay for me to make a mistake.	0	1	2	3	4
16	I can make a good plan and follow through.	0	1	2	3	4
17	I think people at my school care about me.	0	1	2	3	4
18	I will be successful when I grow up.	0	1	2	3	4

*Sample survey instrument altered for one page appendix. Actual GGC instrument formatted as a front/back with larger space to indicate.

Appendix II. GGC Teacher Observation Survey Statements

Instructions:

Steps to Start the Survey

1. Identify the qualified rater: someone who has had regular contact with the student.
2. Fill out rater name, grade, and school.
3. List youth to be rated in column A and provide gender and grade level.
4. "For Time 1, rate based on the first month of school.
5. For Time 2, rate within a month of program end."
6. Answer every question with a number 1-4.
7. Complete fully before moving on to the next student.

Rate based on:

1. Behaviors you have seen (not on what you assume or what you have learned from others).
2. Any interactions during the relevant timeframe
3. The scale only (NOT based on the student's grade level or your perception of the students' overall capacity)
4. Your best understanding (based on observation) of how the student behaved during the relevant timeframe.

Rating Scale

- 1: Never
- 2: Rarely
- 3: Sometimes
- 4: Frequently

Respondent Setup Questions

1. Rater
2. Grade
3. School
4. Gender
5. Grade

To what extent does this student...?

1. Express positive self esteem
2. Define emotional needs
3. Help others without expecting anything in return
4. Get along well with others

5. Understand their actions' impacts on others
6. Express optimism about the world around them
7. Understand and actively participate in learning or positive school culture/climate
8. Understand and believe in his/her ability to make the world a better place
9. Problem solve effectively
10. Work effectively with others in groups

Note: The GGC Survey is distributed to and completed by instructors as a Microsoft Excel File. The instrument has been modified for this format to emphasize construction

Appendix III. Statistical Outputs and Results

All statistical tests performed, certain tests were not included in the report if the test was redundant or underpowered for evaluative purposes per CASEL. Some tests were included descriptively even at non-significance.

A. Student Survey Baseline Equivalence (Pretest)

Group	n	Mean Pre-Score	SD	t(df)	p-value	Cohen's d
Treatment	244	43.57	19.1	-4.24 (360.16)	< .001	-0.40
Control	123	50.19	10.79			

B. Student Survey Within-Group Change (Pre → Post)

Group	n	Mean Pre	Mean Post	Mean Change	t(df)	p-value
Treatment	244	43.57	44.99	+1.43	0.94 (243)	.347
Control	123	50.19	41.07	-9.12	-4.34 (122)	< .001

C. Student Survey Posttest Group Comparison

Group	n	Mean post-score	SD	Mean Difference	t(df)	p-value	Cohen's d
Treatment	244	44.99	20.38	+3.93	1.60 (219.83)	.111	0.18
Control	123	41.07	23.07				

D. ANCOVA Results (Posttest Controlling for Baseline and Grade)

Predictor	B	SE	t	p-value
(Intercept)	26.95	3.96	6.8	< .001
Pretest Score	0.29	0.06	4.52	< .001
Treatment	7.87	2.33	3.37	< .001
Grade 4	2.03	2.5	0.81	.419
Grade 5	-8.9	2.75	-3.23	.001

E. Logistic Regression: Likelihood of Improvement

Predictor	Odds Ratio	p-value
Treatment	1.42	.150

Pretest Score	0.96	< .001
Grade 4	1.67	.051
Grade 5	0.73	.292

F. Baseline Equivalence on Teacher Observation Scores

Outcome	Comparison Mean (SD)	Treatment Mean (SD)	Mean Difference	SMD	p-value
Teacher Observation Score	32.36 (5.58)	32.07 (5.58)	-0.29	-0.05	.51

Note. Baseline equivalence assessed using Welch's two-sample t-test. Standardized mean difference (SMD) calculated using pooled standard deviation.

G. Internal Consistency of the Teacher Observation Survey

Time Point	Items	Cronbach's α
Pre-intervention	10	0.882
Post-intervention	10	0.884

Note. Cronbach's α values indicate good internal consistency.

H. ANCOVA Estimates of Program Effects on Teacher Observation Scores

Predictor	Estimate	Robust SE (HC3)	t	p-value
(Intercept)	1.22	2.89	.422	.673
Treatment	1.94	1.12	1.72	.0849
Baseline Score	.695	.098	7.08	< .0001***
Grade	.976	.452	2.15	.0314*

Note. Models adjust for baseline score and grade level. Robust (HC3) standard errors reported.

I. Teacher Observation Gain Score Model

Predictor	Estimate	Robust SE (HC3)	t	p-value
(Intercept)	-8.001	1.28	-6.24	< .0001***
Treatment	2.01	1.13	1.64	.076
Grade	.673	0.40	1.64	.100

Note. Outcome is change in teacher observation score (post – pre). Robust (HC3) standard errors reported.

Appendix IV. Focus Group Questions

Focus group participants included teachers, administrators, and support staff actively involved in implementing GGC curriculum in Muhlenberg South Elementary School

1. Can you describe your background and experience as an instructor?
2. How did you first learn about GGC's program—what were your initial thoughts about its usage in your school?
3. How would you describe your classroom environment before the program started?
4. What were your first impressions when you began using GGC's program?
5. Can you share any examples of how the program has affected student behavior or engagement towards you as an instructor?
6. Have you noticed any changes in how students work together or interact since starting the program?
7. Have you seen any changes in student behavior outside of formal class time that might be related to the program?
8. How has the program impacted your experience as a teacher?
9. What have you learned about using the program effectively in your classroom?
10. How have your relationships with students changed since using the program?
11. What advice would you give to a teacher who is just starting with the program?
12. Looking back, what's the most important thing you've learned from using the program?

Appendix V. *About the Evaluator, Jarett Lopez*

Jarett Lopez is a 2020 graduate of Northern Kentucky University and a 2022 graduate of the University of Louisville, where he earned a Bachelor of Arts and a Master of Arts in Political Science, respectively. He has professional experience in state government, P–12 and higher education, and nonprofit organizational settings. Jarett is a skilled social scientist with experience in research design, survey development, hypothesis testing, quantitative analysis using R Statistical Computing.

He has previously completed certification in human subjects research through the Institutional Review Boards of Northern Kentucky University and the University of Louisville. His research areas include education and civic development, social identity formation, and the sociopolitical impacts of social media.

Jarett currently manages community outreach, development, and advocacy at the Northern Kentucky office of Kentucky Refugee Ministries in Covington.

Appendix VI. *About Global Game Changers Children’s Education Initiative, Inc.*

Global Game Changers Children’s Education Initiative, Inc. is a national 501(c)(3) nonprofit organization that provides a free, evidence-based Pre-K–5 curriculum focused on social-emotional behaviors, leadership development, and service learning. The program centers on a simple, student-accessible framework—*My Talent + My Heart = My Superpower*—which encourages students to identify their strengths, connect them to causes they care about, and apply those skills in meaningful ways.

Developed through early pilot programs in diverse school settings, the curriculum has since expanded to hundreds of elementary schools nationwide. Global Game Changers is grant funded and made freely available to educators and families through the support of David Novak Leadership, which prioritizes leadership development across the lifespan.

Overview

During the 2024-2025 school year, the Global Game Changers Children's Education Initiative (GGC) implemented a quasi-experimental study in Muhlenberg County Schools. Students at Muhlenberg South Elementary School (MSES) received the GGC curriculum, while students at Bremen Elementary School (BES) continued with typical instruction. Student surveys were administered in fall 2024 (*T1*) and spring 2025 (*T2*).

The main evaluation (Lopez, 2025) found that, after adjustment for baseline differences and grade, students in the intervention school had significantly higher posttest Social Emotional Wellness scores than students in the comparison school.

Follow-up data were collected in spring 2026 (*T3*) from intervention students who remained enrolled at MSES. This supplemental technical appendix examines whether student-reported social-emotional wellness and teacher-observed behavior were maintained or continued to change one year after the original posttest. Because *T3* data were not available for BES, the follow-up analyses describe longitudinal change within the intervention-school cohort and do not estimate a new comparative treatment effect.

Pre/Post GGC Student Survey Data (T1-T3)

The GGC student survey is administered in grades 3-5. At *T3*, current third-grade students were completing the instrument for the first time and therefore had no *T1* or *T2* scores. They were excluded from the longitudinal trajectory analysis, which was limited to current fourth- and fifth-grade students with prior survey opportunities.

Table I presents descriptive statistics for the 93 students who could be matched across all three survey waves. Mean Social Emotional Wellness scores increased from 52.45 at *T1* to 54.61 at *T2* and 55.43 at *T3*, indicating an upward descriptive trajectory in the matched follow-up cohort.

A repeated-measures analysis of variance (*ANOVA*) was used to test whether average scores differed across *T1*, *T2*, and *T3*. A mixed *ANOVA* then examined whether the trajectory varied between grades 4 and 5. Bonferroni-adjusted paired comparisons were used to identify which waves differed after the test.

Across the 107 students with both *T1* and *T3* scores, 64 students (60%) scored higher at follow-up than at baseline. Lower *T1* scores were associated with larger observed changes and greater odds of improvement, while grade 5 was not a significant predictor in either model. Because baseline score is directly related to the amount of possible change, this pattern should be interpreted as baseline-dependent improvement rather than definitive evidence that lower-scoring students experienced a larger causal program effect. The student scale demonstrated good internal consistency at *T1* (Cronbach's $\alpha = .8665$). Bonferroni-adjusted comparisons showed that *T3* scores were significantly higher than *T1* scores by 2.98 points ($p = .023$). The *T1*-to-*T2* difference of 2.16 points was marginal after adjustment ($p = .053$), and the *T2*-to-*T3* difference of 0.82 points was not significant ($p = 1.000$). In context, student-reported wellness was higher at the one-year follow-up than at baseline, while the posttest level was largely maintained rather than followed by a further statistically detectable increase.

The time-only repeated-measures ANOVA indicated a statistically significant change across the three survey waves, $F(2, 184) = 4.995$, $p = .00772$. When grade was included, the wave effect remained significant, $F(2, 182) = 5.097$, $p = .00701$. The overall grade effect was not significant, $F(1, 91) = 1.562$, $p = .215$, and the wave-by-grade interaction did not reach the conventional .05 threshold, $F(2, 182) = 2.891$, $p = .05806$. These results indicate overall change in student-reported wellness without clear evidence that fourth- and fifth-grade students followed different trajectories.

Teacher Observation Survey Data (T1-T3)

Among the 324 students with *T1* and *T3* teacher ratings, 212 students (65%) improved. Lower baseline scores were associated with larger observed gains and greater odds of improvement, although this relationship may partly reflect greater room for growth and regression to the mean. The Teacher Observation Survey demonstrated good internal consistency at *T1* (Cronbach's alpha = .8513).

The mixed ANOVA showed significant effects for grade, $F(4, 259) = 15.62$, $p < .001$, wave, $F(2, 518) = 43.713$, $p < .001$, and the wave-by-grade interaction, $F(8, 518) = 8.658$, $p < .001$. The significant interaction indicates that teacher-observed trajectories were not uniform across grades 1-5. In the focused regression contrast, fifth-grade students showed approximately 2.02 points less *T1*-to-*T3* growth than students in grades 1-4 combined ($p = .0017$) and had lower odds of improvement (OR = .542, $p = .0396$).

Teacher observation scores increased at each wave, from 32.79 at *T1* to 34.05 at *T2* and 35.61 at *T3*. The time-only repeated measures ANOVA was statistically significant, $F(2, 526) = 39.15$, $p < .001$. Bonferroni-adjusted paired comparisons indicated significant increases from *T1* to *T2*, from *T1* to *T3*, and from *T2* to *T3*. Unlike the student self-report pattern, teacher ratings therefore showed continued improvement between the original posttest and the one-year follow-up.

The Teacher Observation Survey consists of 10 items rated from 1 to 4 and summed to a 10-40 scale. The longitudinal teacher analysis included 264 students with complete *T1*, *T2*, and *T3* observations across grades 1-5. Current kindergarten students were new to the follow-up cohort and were not included in the three-wave trajectory analysis.

Discussion and Conclusion

Within the follow-up limitations, the findings support the conclusion that student social-emotional outcomes at MSES did not regress to baseline during the follow-up period. Student-reported wellness was sustained above baseline, and teacher-observed social-emotional and classroom behavior continued to improve. The results are consistent with durable positive development in the intervention-school cohort and support continued implementation accompanied by future data collection that includes a comparison group and grade-specific follow-up.

These results should be interpreted as a longitudinal follow-up of the observed intervention cohort rather than as a new causal evaluation. No comparison-school follow-up data were available, and students included in the three-wave analyses were those who remained enrolled and could be matched over time. Maturation, school context, staff practices, and selective follow-up may therefore contribute to the observed patterns.

The grade findings also differ by measure. Student survey trajectories were generally similar for grades 4 and 5. Teacher-observed trajectories varied significantly across grades, and fifth-grade students showed weaker improvement than students in grades 1-4 combined. This may reflect developmental differences, changing classroom expectations, variation in implementation, or differences in how teachers apply the rating scale. Grade-specific means would be needed to identify the precise source of the interaction.

Convergence across measures is substantively important. Both sources showed higher T3 scores than T1 scores, and a majority of students improved from baseline to follow-up. At the same time, the measures captured different aspects of student functioning. Student self-report reflects perceptions of belonging, emotional competence, optimism, and personal agency, whereas teacher ratings emphasize visible classroom behavior such as cooperation, problem solving, participation, and helping others. The stronger T2-to-T3 change in teacher ratings may indicate that observable behaviors continued to consolidate even after students' self-perceptions had stabilized.

Taken together, the student and teacher data indicate that social-emotional outcomes in the intervention school remained above baseline one year after the original posttest. Student self-reports suggest that gains were primarily maintained: the T3 mean was significantly higher than T1, but T2 and T3 did not differ. Teacher observations showed a stronger pattern of continued development, with statistically significant increases across each interval.

Statistical Outputs

Table I. Student Survey - Three-Wave Descriptive Statistics

Wave	n	Mean	SD	Minimum	Maximum
T1 Pre	93	52.45	10.382	31	72
T2 Post	93	54.61	8.759	34	72
T3 Follow-up	93	55.43	8.306	36	71

Table II. Student Survey - Repeated-Measures ANOVA

B1. Time-Only Model

Effect	df	Residual df	Sum of Squares	Mean Square	F	p-value
Wave	2	184	441	220.3	4.995	.00772

B2. Mixed ANOVA: Wave x Grade

Effect	df	Residual df	Sum of Squares	Mean Square	F	p-value
Grade	1	91	257	256.7	1.562	.215
Wave	2	182	441	220.26	5.097	.00701
Wave x Grade	2	182	250	124.93	2.891	.05806

Table III. Student Survey - Bonferroni-Adjusted Pairwise Comparisons

Comparison	Earlier Mean	Later Mean	Mean Difference	Adjusted p-value
T1 vs. T2	52.45	54.61	+2.16	.053
T1 vs. T3	52.45	55.43	+2.98	.023
T2 vs. T3	54.61	55.43	+0.82	1.000

Table IV. Student Survey - Linear Regression of T1-to-T3 Change

Model: (T3 - T1) ~ T1 + Grade 5

Predictor	B	SE	t	p-value
Intercept	40.00152	3.76010	10.638	< .001
T1 score	-0.71034	0.06745	-10.532	< .001
Grade 5	1.03753	1.47699	0.702	.484
Model statistic		Value		
Analytic n		107		
Residual df		104		
Residual standard error		7.522		
R ²		.5306		
Adjusted R ²		.5215		
Model F		58.77		
Model df		2, 104		
Model p-value		< .001		

Table V. Student Survey - Logistic Regression Predicting Improvement

Outcome: T3 score greater than T1 score

Predictor	Log-Odds B	SE	z	p-value	Odds Ratio
Intercept	10.68388	2.25235	4.743	< .001	43,646.76

Predictor	Log-Odds B	SE	z	p-value	Odds Ratio
T1 score	-0.19776	0.04086	-4.841	< .001	0.821
Grade 5	0.65424	0.51901	1.261	.207	1.924
Model statistic			Value		
Analytic n			107		
Improved			64		
Percent improved			60%		
Null deviance			144.19		
Residual deviance			91.46		
AIC			97.46		

Table VI. Student Survey - Reliability

Scale	Wave	Cronbach's alpha
Social Emotional Wellness	T1	.8665

Table VII. Teacher Observation Survey - Three-Wave Descriptive Statistics

Wave	n	Mean	SD	Minimum	Maximum
T1 Pre	264	32.79	5.374	14	40
T2 Post	264	34.05	5.046	16	40
T3 Follow-up	264	35.61	5.008	19	40

Scale: Ten items rated 1-4 and summed as-is, yielding a possible range of 10-40.

Table VII. Teacher Observation Survey - Repeated-Measures ANOVA

H1. Time-Only Model

Effect	df	Residual df	Sum of Squares	Mean Square	F	p-value
Wave	2	526	1,052	526.0	39.15	< .001

Effect	df	Residual df	Sum of Squares	Mean Square	F	p-value

H2. Mixed ANOVA: Wave x Grade

Effect	df	Residual df	Sum of Squares	Mean Square	F	p-value
Grade	4	259	2,686	671.5	15.62	< .001
Wave	2	518	1,052	526.0	43.713	< .001
Wave x Grade	8	518	833	104.2	8.658	< .001

Table VIII. Teacher Observation Survey - Bonferroni-Adjusted Pairwise Comparisons

Comparison	Earlier Mean	Later Mean	Mean Difference	Adjusted p-value
T1 vs. T2	32.79	34.05	+1.26	1.6×10^{-7}
T1 vs. T3	32.79	35.61	+2.82	6.4×10^{-13}
T2 vs. T3	34.05	35.61	+1.56	3.8×10^{-5}

Table IX. Teacher Observation Survey - Linear Regression of T1-to-T3 Change

Model: $(T3 - T1) \sim T1 + \text{Grade } 5$

Predictor	B	SE	t	p-value
Intercept	24.68404	1.68424	14.656	< .001
T1 score	-0.65975	0.05158	-12.791	< .001
Grade 5	-2.01650	0.63713	-3.165	.0017
Model statistic		Value		
Analytic n		324		
Residual df		321		
Residual standard error		4.816		
R ²		.3628		

Predictor	B	SE	t	p-value
Adjusted R ²			.3588	
Model F			91.39	
Model df			2, 321	
Model p-value			< .001	

Table X. Teacher Observation Survey - Logistic Regression Predicting Improvement

Outcome: T3 score greater than T1 score

Predictor	Log-Odds B	SE	z	p-value	Odds Ratio
Intercept	8.18324	1.09966	7.442	< .001	3,580.45
T1 score	-0.22188	0.03185	-6.965	< .001	0.801
Grade 5	-0.61236	0.29759	-2.058	.0396	0.542
Model statistic			Value		
Analytic n			324		
Improved			212		
Percent improved			65%		
Null deviance			417.79		
Residual deviance			347.22		
AIC			353.22		

Table XI. Teacher Observation Survey - Reliability

Scale	Wave	Cronbach's alpha
Teacher Observation Survey	T1	.8513