CharacterStrong | RESEARCH BRIEF

Program Evaluation Report of the CharacterStrong Middle/High School (Grades 6-12) Life Skills Program

Yanchen Zhang, PhD, NCSP
Assistant Professor of Psychological & Quantitative Foundations, University of Iowa

Clayton Cook, PhD, LP
Professor, John W. and Nancy E. Peyton Faculty Fellow in Child and Adolescent Wellbeing, Department of Educational Psychology, University of Minnesota

Brian Smith, PhD
Director of Research, CharacterStrong

Table of Contents

Overview

This report contains the details and primary outcomes of a program evaluation study for the CharacterStrong Secondary School (Grades 6-12) Life Skills Program (CS curriculum), which was conducted in a mid-sized public school district in the Pacific Northwest. In a cluster randomized controlled trial, fourteen secondary (middle and high: grades 6-12) schools were randomly assigned to an intervention or control condition. Teachers in intervention schools were supported to deliver the CS curriculum on a weekly basis. The study used online surveys to collect pre- and post-test data from students and teachers as well as gathered administrative data on disciplinary referrals and attendance. Findings indicated significant effects of CS the curriculum on outcomes gathered from students, teachers, and school administrative records. Specifically, this study found (a) students in the intervention condition showed higher levels of engagement behaviors conducive to learning and perceived school safety than students in the control condition, (b) teachers reported significant higher levels of safety in intervention schools, and (c) school administrative data indicated significantly fewer incidents of unexcused absences and truancies in the intervention schools.

Design and Procedures

Study Design

We systematically evaluated the effectiveness of the school-based implementation of the CharacterStrong (CS) curriculum using both process- and outcome-focused evaluation. Based on the gold standard in the literature of program evaluation (Dowling, Simpkin, & Barry, 2019), we adopted cluster randomized controlled trials (pre-post randomized controlled trials) as the primary design of the quantitative evaluation of the effectiveness of the CS curriculum at the middle and high school level. In August 2019, a total of 14 secondary schools (n student = 7113, n teacher = 519) located in the Pacific Northwest were first matched into pairs based on demographics. Then each school from a pair was randomly assigned to either the intervention group (i.e., implementing the CS curriculum at the universal level during the regular school day) or the control group (i.e., doing regular school activities as usual without receiving any intervention components from the CS curriculum). The use of school-level matching and random assignment was to (a) optimize the probability that the two groups were equivalent in terms of school demographics and outcome variables (e.g., climate, safety, and sense of belonging) at baseline, and (b) preclude the potential group contamination effect where participants from the intervention or control groups unintentionally influence each other because they are from the same school building (Dowling et al., 2019; Hemming, & Girling, 2019).

To adequately measure the growth associated with the school-wide implementation of the CS curriculum, we collected pre-and post-test data about key student and teacher outcomes (Student: school climate and attributes, Teacher: school climate) before and after the implementation of the CS curriculum. To facilitate pre-/post-test data collection, a series of standard web-based surveys were constructed for students and teachers, specifically, by strategic selection and adoption of scales from the Panorama Life Skills Student Survey by the CS research team in consultation with the participating schools (Panorama Education, 2016; see the Measures section).

Recruitment

The CS research team partnered with the school districts' research and evaluation department in recruitment and program evaluation. The initial selection of schools involved working with central administrators and communicating with site-based administrators regarding the project's benefits and data collection procedures. Schools were recruited if they were interested in implementing the CS curriculum as their life skills initiative for universal prevention and promotion. Once a pool of candidate schools was identified, they were randomly selected for participation. This evaluation study rolled out in multiple steps from September 2019 to June 2020 (Figure 1). In October 2019, our CS research team collaborated with the school district research and evaluation departments to collect the pre-test data using the student and teacher versions of web-based surveys about the target outcomes from all participating schools. Specifically, an initial email was sent by the CS research team to all participating schools to provide them with an overview of the project, obtain informed consent, and provide a link to the online survey. Each school was provided with a one-month window to complete the survey from the time they were sent the initial email. Reminder emails were sent weekly to increase the number of respondents from each school.

After pre-test completion, schools in the intervention group received the CS curriculum while the schools in the control group continued with business as usual. When the implementation of the CS curriculum was finished for the intervention group in May 2020, the post-test data were collected from all participating schools using the same web-based surveys (student and teacher versions). The multi-step intervention and data collection procedure controlled for the temporal sequencing among effects of the implementation of the CS curriculum, and outcomes at baseline and post-intervention (i.e., pre-/post-tests). Our CS research team implemented an internal ethics process audited by the leadership and research and evaluation departments of the participating school district, which provided all participants of the CS curriculum with an Information Sheet and Permission Form to gain their informed consent to collect data for research and evaluation purposes. There was no distinguishable data that could be connected to specific students as a part of this study. All students were given random ID codes by the district so as to protect privacy. This study was considered exempt from IRB because it was part of a program evaluation by the school district and CharacterStrong and none of the data

gathered were outside of the scope of information that the school district normally gathers as part of routine practice.

Training

Full-day in-person training on the curriculum was provided by CharacterStrong staff to intervention school staff.

Participants

As described in the Design section, we initially recruited for program evaluation a total of 14 secondary schools (n student = 7113, n teacher = 519) from a large and diverse urban school district in the Northwest. The pre- and post-test data were collected with three versions of surveys (two for students, one for teachers), therefore, the final analytic samples of participants are summarized in three sub-sections. Of note, the covid-19-19 pandemic in Spring 2020 led to massive school closure and change of teaching modality in the participating schools. Therefore, we observed a large attrition rate in the post-test data collection. The method used to handle missingness in the final analytic samples is reported in the Analysis Plan section. As an effort to optimize the accuracy and transparency in reporting, we present the descriptive statistics for both the initial sample and the final analytic sample used to evaluate the effectiveness of the CS program (i.e., remaining cases after list-wise deletion of all the cases missing on pretest outcomes). Table 1 is a summary of the numbers of participants in each study group at pre- and Post-test.

School-level Demographics

Fourteen secondary schools participated in this program evaluation (n school = 14, n teacher = 519, n students = 7,113), from a large and diverse urban public school district on the West Coast. Participating schools were racially/ethnically (M Percentage of Non-White = 30.3%; minimum = 25%, maximum =45.90 %) and socioeconomically (e.g., M Percentage of free and reduced-priced lunch = 55.99 %; minimum = 22.2%, maximum = 83.9%) diverse. For details of School-level demographics in the dataset, see Table 2.1.

Initial Sample for Student Surveys

In the initial sample at the pretest, 50.8% of the students identified as male. Most students (69.7%) identified their race/ethnicity as Caucasian, followed by 12.8% Multiracial, 10.1% Hispanic, and 2.96% Black or African American. About half of respondents (48.57%) received FRPL (Table 2.2). Due to missing, the percentage may not correspond exactly to the school-level demographics.

Final Sample for Student Attribute Survey

Due to COVID-19-19 pandemic, about 83.31% cases from pre-test (npretest = 6329) were missing at the post-test in Spring 2020 (npost = 1056). In the final analytic sample, 37.4% of the students identified as male. Most students (71.3%) identified their race/ethnicity as Caucasian, followed by 10.6% Multiracial, 9.9% Hispanic, and 4.6% Asian. See Table 2.3.

Final Sample for Student School Climate Survey

Due to the COVID-19 pandemic, about 85.11% of cases from pre-test of the Student School Climate Survey (npretest = 7112) were missing at the post-test in Spring 2020 (nposttest = 1059). In the final analytic sample, 37% of the students identified as male. Most students (72.2%) identified their race/ethnicity as Caucasian, followed by 10% Multiracial, 9.3% Hispanic, and 4.6% Asian. See Table 2.4.

Final Sample for Teacher School Climate Survey

Due to COVID-19 pandemic, about 39.31% of cases from pre-test of the Student School Climate Survey (npretest = 519) were missing at the post-test in Spring 2020 (nposttest = 315; see Table 4.3). About 42.5% of educators surveyed self-identified as male.

Measures

Student & Teacher Demographics

A set of student and teacher demographic variables were added to the survey data (Table 2.2), which are theoretically relevant to the research questions and commonly used in educational or implementation research. The demographic variables were obtained via either administrative data reported by the participating schools or the self-report data from the first section of the web-based survey.

The demographic variables collected include student/teacher gender, racial ethnicity, and status in the English Language Learning, Special Education, and Free-/Reduced-Priced Lunch programs. To eliminate potential confounding effects and optimize statistical power, the demographic covariates would be included in the analytic models if statistical equivalence between intervention and control groups was not established at baseline (pre-test).
Surveys built from the Panorama Life Skills Student Survey

In consultation with the participating schools, the CS research team worked with Panorama Education to strategically selected scales from the Panorama Life Skills Student Survey (P-LSSS; Panorama Education, 2016). Adopted by over 1,000 schools, the P-LSSS is a comprehensive suite of multiple validated scales to assess the outcomes associated with the implementation of school-based Life Skills curricula, such as the CS curriculum. The whole suite targets capturing the change in three crucial domains related to student Life Skills outcomes, social relationships (Gehlbach et al., 2016; Walton & Cohen, 2011), motivation (Hulleman & Harackiewicz, 2009), and self-regulation (Duckworth, Kirby, Tsukayama, Berstein, & Ericsson, 2011). Scales of the P- SELS include student self-reports on personal attributes, student perceptions of the school environment and supports, and teachers' perspectives and skills. In a large-scale validation study of the P-LSSS, the scales were found to be highly reliable and of acceptable validity (Panorama Education, 2020).

To facilitate pre-/post-test data collection, three web-based surveys were constructed specifically for students (Student Attribute Survey and Student School Climate Survey) and teachers (Teacher School Climate Survey). These surveys were categorized by the seven scales adopted from the P-LSSS considering its key features and relevance to the learning goals of the CS curriculum. For instance, the Student Attribute Survey consists of scales of the P-LSSS focusing more on the individual student (e.g., grit, mindset, Self-Efficacy), while the Student School Climate Survey adopted scales of the P-LSSS that focus more on the environment (e.g., classroom climate and School Safety). The following section details the content and psychometrics of the scales adopted from the P-LSSS into the three surveys.

Scales in the Student Attribute Survey

We assessed students' personal attributes before and after the implementation of the CS curriculum using seven scales from the P-LSSS: Grit (item number =5), Learning Strategies (item number =5), Self-Efficacy (item number =5), Self-Management (item number =10), Social Awareness (item number =8), and Growth Mindset (item number =6) (Panorama Education, 2020). Students indicated the extent to which they agreed with the statements in each measure using a 5-point Likert scale ranging from 1 to 5. The specific label for each Likert-scale point varied by specific scales in accordance with the current best practices in the science of survey design. A sample item from the Self-Efficacy scale is "How confident are you that you can complete all the work that is assigned in your classes?", where the student is asked to rate the extent to which they agreed with the statements in each measure using a 5-point Likert-scale

ranging from 1 " Not at all Confident" to 5 " Extremely Confident". To calculate scale scores, we recoded negatively worded items to ensure that 5 always reflected a more positive response, then averaged across item scores within each scale. The pre-test was completed between October 1 and November 8, 2019, while the post-test was completed between May 25 and Jun 19, 2020. The reliability estimates for each scale range from acceptable to high (Table 3).
Of note, we added to the P-LSSS an ad hoc subscale, Learning Behaviors (item # = 5), to capture changes in the frequency of students' specific behaviors conducive to learning at school, which is an important target in the design of the CS curriculum. The subscale consists of five items:

1. I show up to class on time.
2. I complete and turn in class assignments on time.
3. I participate in classroom discussions.
4. I exhibit behaviors that are disruptive to learning. (reversed)
5. I try my best to be a good citizen of this school.

The student is asked to rate the extent to which they agreed with the statements in each measure using a 5-point Likert scale ranging from 1 "None of the time" to 5 "All of the time". The reliability estimate of the Learning Behaviors is acceptable based on results from the final analytic sample (Cronbach's alpha = .58; Table 3).

School Climate Surveys

Considering that a target goal of the CS curriculum is to create a safe and nurturing learning environment at the school level, we assessed various aspects of school climate from the perspectives of both students and teachers. The reliability estimates for each scale range from acceptable to high (Table 3).

Scales in the Student School Climate Survey

As the primary beneficiaries of the CS curriculum, we assessed students' perceptions of constructs related to school climate before and after the implementation of the CS curriculum using seven scales from the P-LSSS, including Engagement (item number =5), Rigorous Expectations (item number =5), School Climate (item number =5), School Safety (item number =6), Sense of Belonging (item number =5), Teacher-Student Relationships (item number =5), Valuing of School (item number =5; Panorama Education, 2020).

In each scale, students indicated the extent to which they agreed with the statements in each measure using a 5-point Likert- scale ranging from 1 to 5. The specific label for each Likert-scale point varied by specific scales in accordance with the current best practices in the science of survey design. A sample item from the School Safety scale is “How often do you worry about violence at your school?”, where the student is asked to rate the extent to which they agreed with the statements in each measure using a 5-point Likert-scale ranging from 1 "Almost never" to 5 " Almost always". To calculate scale scores, we first recoded negatively worded items (e.g., the example just given from the School Safety scale) to ensure that 5 always reflected a more positive response, then averaged across item scores within each scale. The pre-test was completed between October 1 and November 8, 2019, while the post-test was completed between May 25 and Jun 19, 2020.

Scales in the Teacher School Climate Survey

We assessed teachers perceptions of constructs related to school climate before and after the implementation of the CS curriculum using seven scales from the P-LSSS, including School Climate-educator version (item number =9), Professional Learning (item number =8), Feedback and Coaching (item number =5), School Leadership (item number =9), Educating All Students (item number =9), Teacher Self-Reflection (item number =8), and School Safety- educator version (item number =3; Panorama Education, 2020).

In each scale, teachers indicated the extent to which they agreed with the statements in each measure using a 5-point Likert- scale ranging from 1 to 5. The specific label for each Likert-scale point varied by specific scales following the current best practices in the science of survey design. A sample item from the Professional Learning scale is “Overall, how supportive has the school been of your growth as a teacher?”, where the teacher is asked to rate the extent to which they agreed with the statements in each measure using a 5-point Likert-scale ranging from 1 "Not at all supportive" to 5 " Extremely supportive". To calculate scale scores, we recoded negatively worded items to ensure that 5 always reflecting a more positive response, then averaged across item scores within each scale. The pre-test was completed between October 15 and November 8, 2019, while the post-test was completed between May 25 and Jun 19, 2020.

Administrative Student Behavioral Data

To corroborate the primary outcomes (i.e., scales from P-LSSS), we collected end-of-year (June 2020) administrative behavioral data from participating schools as ancillary outcomes. There were three types of administrative data collected and matched to the student participants based on their ID numbers, including frequency of serious misconduct reports, and truancy (two types of Average Cumulative Absence rate - ACA-all and ACA- unexcused/truancy only) calculated based on the entire 2019-2020 academic year (Table 7).

The ACA-all coding is calculated by dividing the number of periods missed for any reason by the number of potential periods in the academic year. The ACA- unexcused/truancy only is calculated by dividing the number of periods missed due to unexcused absence or truancy by the number of all potential periods. We performed log transformation with a base of 10 to rectify the left skewness of the distributions of the two types of ACA, which resulted in an approximately normal distribution. Given the percentage nature of the original ACA variables, we used equal percentage binning techniques to bin the log-transformed ACA variables into categorical ones with five levels to facilitate analysis and result interpretation (5= extremely high percentage of absence, 4 = high, 3 = moderate, 2= low, 1= extremely low).

Analysis Plan

Missing Data Handling

A preliminary missing data analysis revealed that the current dataset consisted of a large proportion of missing cases from pre- test to post-test due to the significant disruptions to schooling due to the COVID-19-19 pandemic in Spring 2020 (Table 1). All data analytic approaches in this evaluation study employed the list-wise deletion method (Dempster et al., 1977) to adjust for any missing data and reduce potential biases. This is because the list-wise deletion method is relatively less biased compared to pair-wise deletion (the alternate option) while allowing for less biased estimates of regression coefficients and standard errors under the assumption that the missing data are missing at random (Schafer, 1997), which is realistic given that every precaution was made to ensure that participant non-response was not associated with their inclusion or omission in the study (e.g., examine patterns of missing data for potential biases in sampling design or participant recruitment).

Baseline Equivalence

With pre-test data in the final analytic sample (i.e., the cases with both pre and post-tests), we established the baseline equivalences of demographic variables (Chi-square tests) and outcomes (independent t-tests between intervention and control groups).

Demographics

Table 4 presents detailed information about the statistical baseline equivalences of the demographic covariates between the intervention and control groups in the final analytic samples across the three surveys. For the Student Attribute Survey, Student and Teacher School Climate Surveys, the results of the Chi-square tests demonstrated statistical baseline equivalence across all demographic variables. To optimize the statistical power and abide by

the principle of parsimony no demographic variables were entered into the final analytic models because of their equivalence between intervention and control groups.

Primary Outcomes at Pre-Test

Table 5 presents the results from independent t-tests detecting the statistical baseline equivalences of primary outcomes (i.e., scales from the P-LSSS) in the final analytic samples across the three surveys. For all primary outcomes, there are no statistical differences between the intervention and control groups. The established baseline equivalence in the participants' demographics and primary outcomes at pretest was expected because of the use of matching and random assignment in the design of this evaluation study.

Program Effectiveness Tests

Although ANCOVA (i.e., controlling outcome at pre-test as covariates) or repeated measure ANOVA are the best practices in analyzing data obtained from pre-post randomized controlled trials, the massive missing cases between pre-and post-tests rendered these two approaches impractical (Wan, 2020). Furthermore, due to the small number of the clusters randomized and the massive missingness in this study, we did not have enough statistical power to perform multilevel modeling.

We originally planned to use the approach of "Intention-to-Treat" (ITT; Gupta, 2011) to handle missing cases from pre-test to post test, but it turned out to be impractical and unreasonable given the significant disruption to the study from COVID-19-related school closures. First, considering the unique situation of this evaluation study, we randomly recruited extra participants to both groups to ensure the balance and equivalence in demographics to make up, to a certain extent, for the disproportionate cases between intervention and control groups due to missingness. Second, theoretically, the massive missingness de facto resulted in the change of our study design from a pre-posttest controlled trial to a posttest-only randomized experimental design. It is unlikely any significant bias was introduced by initial differences between the intervention and control groups considering that the baseline equivalence was established in the primary outcome variables (P-LSSS) and demographics at pre-test. Furthermore, if we use best practices in ITT (e.g., imputation or carry-over of pre-test), the massive missingness from pre to post-tests due to the COVID-19 pandemic will force us to impute disproportionately large amounts of cases/values which were merely statistical artifacts. This will result in largely biased estimation of the program effects, which is not generalizable to real-world observations and will preclude the potential to detect any statistically or clinically meaningful results from this large-scale evaluation study.

To optimize analytic integrity, we conducted the intervention effect analysis with the post-test data only, while using independent sample t-tests (with continuous outcome variables) and Chi-square test of independence (with categorical outcome variables). This approach ensured that the estimates of the intervention effect of the CS curriculum are robust and not biased due to the extremely unbalanced pre versus post-test outcomes caused by the missingness (Dugard & Todman, 1995). All analyses were performed with SPSS version 24 (IBM Corp. Armonk, NY, 2018). First, we ran descriptive statistics (e.g., measures of central tendency and variability, heteroskedasticity, and variable distributions) for outcomes. The descriptive statistics were further checked for evidence of the adequacy of the data for independent sample t-test (e.g., comparable scales of measurement, normality, and homoscedasticity). To detect potential significant differences between the student and teacher outcomes in the intervention group compared to those in the comparison group, we ran independent sample t-tests for all primary outcome variables and Chi-square test of independence using the binary-coded group variable (intervention group =1; control group = 0) with the three datasets from the Student Attribute Survey, Student School Climate Survey, and Teacher School Climate Survey.

Results

Baseline Equivalency Testing

For the results of baseline equivalency in pre-test data and demographic variables, please see the Analysis Plan section. In sum, we established baseline equivalences for all demographic variables via Chi-square tests (Table 4) and primary outcomes at pre-test (independent t-tests between intervention and control groups (Table 5; Figures 1 to 3).

Treatment Effect Analysis

The results from the three surveys at post-test revealed some significant findings in selected primary and ancillary outcomes in favor of the intervention group compared to the control group (Table 6; Figures 4 to 6).

In the Student Attribute Survey, we identified that the students in the intervention group demonstrated more behaviors conducive to promoting learning at school compared to the students in the control group. The scale score of Learning Behaviors in the intervention group ($M = 4.069$; $SD = 0.564$) was higher than the control group ($M = 3.999$; $SD = 0.56$). The independent sample t-test result was statistically significant, $t\ (1054) = 1.977$, $p < .05$. The effect size, as in Hedges' g, is 0.125 (Ellis, 2010).

For the Student School Climate Survey, we found a significantly higher level of perceived school safety reported by the students in the intervention group compared to those in the control

group. The scale score of School Safety in the intervention group (M = 4.125; SD = 0.669) was higher than the control group (M = 3.956; SD = 0.68). The independent sample t-test result was statistically significant, t (1056) = 3.942, p < .001. The effect size, as in Hedges' g, is 0.25. The responses from the Teacher School Climate Survey showed convergent results with the student version at post-test.

Similar to the students, we found a significantly higher level of perceived school safety reported by the teachers in the intervention group compared to those in the control group. The scale score of School Safety-Educator version in the intervention group (M = 3.807;

SD = 0.843) were higher than the control group (M = 3.46; SD = 0.88). The independent sample t-test result was statistically significant, t (313) = 2.715, p < .01. The effect size, as in Hedges' g, is 0.397.

The other subscales demonstrated certain differences in their mean scale scores, with some in favor of the intervention group as compared to the control. But their mean differences do not approach statistical significance. As to the ancillary outcomes, the results from the Chi-square test of independence indicated that the students in the intervention group had significantly fewer incidents of unexcused absences or truancies as compared to those in the control group, $\chi 2$ (4, n =571) = 16.161, p < .01. The results for frequency of serious misbehavior and general absence did not reach statistical significance (Table 7).

Bibliography:

1. Dowling, K., Simpkin, A. J., & Barry, M. M. (2019). A cluster randomized-controlled trial of the mindout learning program for disadvantaged post-primary school students. Journal of Youth and Adolescence, 48(7), 1245-1263.
2. Hemming, K., & Girling, A. (2019). Cluster randomised trials: useful for interventions delivered to groups: Study design: cluster randomised trials. BJOG: An International Journal of Obstetrics & Gynaecology, 126(3), 340-340.
3. Panorama Education (2016). Validity Brief: Panorama Student Survey. https://go.panoramaed.com/hubfs/Panorama_January2019%20/Docs/validity-brief.pdf
4. Panorama research team. (2020). Reliability and Validity of Panorama Survey Topics for Students: 2024 Update.https://www.panoramaed.com/hubfs/Validity-Report-Student-Topics-2024.pdf
5. Wan, F. (2020). Statistical analysis of two arm randomized pre-post design with one post-treatment measurement. arXiv preprint arXiv:2007.07881.

Table 1

Descriptive Statistics of the Analytic Samples of Student at Pre-Post Tests

Survey Type	Group	Pre-Test	Post-Test	Attrition

Table 2.1

Descriptive Statistics of School Demographics

	N	Range	Minimum	Maximum	Mean	Std. Deviation
%Male student	14	15.20	44.10	59.30	50.56	3.76
%Non-White	14	20.90	25.00	45.90	33.43	6.97
%FRPL	14	61.70	22.20	83.90	55.99	21.56
%ELL	14	19.60	3.40	23.00	10.05	6.16
%Special Education	14	22.70	0.00	22.70	10.44	4.98

Table 2.2

Descriptive Statistics of Initial Sample for Students at Pretest

		Frequency	Percent	Valid Percent
Grade Level	7	1373	21.69	21.69
	8	1331	21.03	21.03
	9	1063	16.80	16.80
	10	982	15.52	15.52
	11	813	12.85	12.85
	12	767	12.12	12.12
	Total	6329	100	100
Gender	Female	3113	49.19	49.21
	Male	3213	50.77	50.79
	Missing	3	0.05	
	Total	6329	100	
Race	Caucasian	4410	69.68	69.69
	African-American	187	2.95	2.96
	Asian	151	2.39	2.39
	Hispanic	639	10.10	10.10
	Native American	46	0.73	0.73
	Pacific Islander	84	1.33	1.33
	Multi-Racial	811	12.81	12.82
	Missing	1	0.02	
	Total	6329	100	
FRPL	Paid	3019	47.70	49.55
	Reduced	687	10.85	11.28
	Free	2387	37.72	39.18
	Missing	236	3.73	
	Total	6329	100	
ELL	No	5516	87.15	90.52
	Yes	578	9.13	9.48
	Missing	235	3.71	
	Total	6329	100	
Special Education	No	5479	86.57	89.91
	Yes	615	9.72	10.09
	Missing	235	3.71	
	Total	6329	100	

Table 2.3

Descriptive Statistics of Final Sample for Student Attribute Survey

		Frequency	Percent
Grade	6	28	2.7
	7	187	17.7
	8	201	19.0
	9	165	15.6
	10	176	16.7
	11	139	13.2
	12	160	15.2
	Total	1056	100
Gender	Female	661	62.6
	Male	395	37.4
	Total	1056	100
Race	Caucasian	753	71.3
	African-American	24	2.3
	Asian	49	4.6
	Hispanic	105	9.9
	Native American	10	0.9
	Pacific Islander	3	0.3
	Multi-Racial	112	10.6
	Total	1056	100
FRPL	Paid	619	58.6
	Reduced	88	8.3
	Free	349	33.0
	Total	1056	100
ELL	No	1017	96.3
	Yes	39	3.7
	Total	1056	100
Special Education	No	987	93.5
	Yes	69	6.5
	Total	1056	100

Table 2.4

Descriptive Statistics of Final Sample for Student School Climate Survey

		Frequency	Percent	Valid Percent
Grade Level	6	28	2.60	2.60
	7	175	16.50	16.50
	8	190	17.90	17.90
	9	183	17.30	17.30
	10	183	17.30	17.20
	11	140	13.20	13.20
	12	160	15.10	15.10
	Total	1059	100	100
Gender	Female	667	63.00	63.0
	Male	391	36.90	37.0
	Total	1058	99.90	100.0
	Missing	1	0.10	
	Total	1059	100	
Race	Caucasian	765	72.2	72.2
	African-American	26	2.5	2.5
	Asian	49	4.6	4.6
	Hispanic	99	9.3	19.3
	Native American	10	0.9	0.9
	Pacific Islander	4	0.4	0.4
	Multi-Racial	106	10.0	10.0
	Total	1059	100	
FRPL	Paid	634	59.9	59.9
	Reduced	85	8.0	8.0
	Free	340	32.1	32.1
	Total	1059	100	100
ELL	No	1024	96.7	96.7
	Yes	35	3.3	3.3
	Total	1059	100	100
Special Education	No	995	94.0	94.0
	Yes	64	6.0	6.0
	Total	1059	100	100

Table 2.5

Descriptive Statistics of Final Sample for Teacher School Climate Survey

		Frequency	Percent
Initial sample	Female	290	55.88
	Male	229	44.12
	Total	519	100
Final sample	Female	181	57.5
	Male	134	42.5
	Total	315	100

Table 3

Reliability Estimates by Scales

Scales	a
Student Attribute Survey	
Grit	.74
Learning Strategies	.83
Learning Behaviors*	.58
Self-Efficacy	.78
Self-Management	.83
Social Awareness	.80
Growth Mindset	.79
Student School Climate Survey	
Engagement	.87
Rigorous Expectations	.86
School Climate	.84
School Safety	.73
Sense of Belonging	.85
Teacher-Student Relationships	.90
Valuing of School*	.81
Teacher School Climate Survey	
School Climate-educator version*	.85
Professional Learning	.89
Feedback and Coaching*	.91
School Leadership*	.93
Educating All Students*	.89
Teacher Self-Reflection	.94
School Safety- educator version*	.64

Note. α = Cronbach's alpha, * indicate the reliability coefficient is estimated using our final analytic sample.

Table 4

Chi-Square tests for statistical equivalence of demographic covariates in the final sample

Survey	Group	χ^2	*df*	*p*
Student Attribute	Grade Level	8.047	5	.154
	Gender	0.013	1	.909
	Race	2.126	1	.145
	FRPL	2.861	2	.239
	ELL	1.928	1	.165
	Special Education	0.834	1	.361
Student School Climate	Grade Level	6.204	5	.196
	Gender	0.268	1	.605
	Race	2.28	1	.131
	FRPL	0.608	2	.738
	ELL	0.594	1	.441
	Special Education	0.887	1	.346
Teacher School Climate	Gender	0.495	1	.482

Note. Due to extremely small frequencies of some racial groups, the race variable was recoded to be binary: 0 = White/Caucasian, 1 = Non-White.

Table 5

Independent t-tests of Outcome Variables Between the Treatment and Control Groups at Pre-Test

Survey	Outcomes at Pre-Test	Grp	*n*	*Mean*	*SD*	*t*	*df*	*p*	Mean Difference	Std. Error	95% CI of Mean Difference	
											Lower	Upper
Student Attribute	Grit	Intv	3840	3.435	0.721	0.143	7106	.886	0.002	0.017	-0.031	0.036
		Ctrl	3268	3.433	0.713							
	Growth Mindset	Intv	3829	3.511	0.797	0.649	7085	.517	0.012	0.019	-0.025	0.049
		Ctrl	3258	3.499	0.781							
	Learning Behaviors	Intv	3821	3.867	0.611	0.417	7065	.677	0.006	0.015	-0.023	0.035
		Ctrl	3246	3.861	0.630							
	Learning Strategies	Intv	3819	3.468	0.768	1.257	7061	.209	0.023	0.018	-0.013	0.059
		Ctrl	3244	3.445	0.771							
	Self-Efficacy	Intv	3834	3.251	0.810	0.308	7095	.758	0.006	0.019	-0.032	0.044
		Ctrl	3263	3.245	0.812							
	Self-Management	Intv	3830	3.902	0.630	1.904	7089	.057	0.029	0.015	-0.001	0.058
		Ctrl	3261	3.874	0.629							
	Social Awareness	Intv	3832	3.580	0.646	0.225	7092	.822	0.004	0.016	-0.027	0.034
		Ctrl	3262	3.576	0.664							

Student Climate	Engagement	Intv	3787	2.911	0.839	1.199	6320	.230	0.026	0.022	-0.016	0.068
		Ctrl	2535	2.885	0.849							
	Rigorous Expectations	Intv	3785	3.625	0.719	0.500	6316	.617	-0.009	0.018	-0.045	0.027
		Ctrl	2533	3.634	0.708							
	School Climate	Intv	3785	4.246	1.086	1.328	6318	.184	0.037	0.028	-0.018	0.093
		Ctrl	2535	4.209	1.111							
	School Safety	Intv	3771	3.461	0.773	1.456	5415	.145	0.046	0.020	0.006	0.086
		Ctrl	2521	3.415	0.819							
	Sense of Belonging	Intv	3770	2.987	0.824	1.455	6288	.146	0.031	0.021	-0.011	0.072
		Ctrl	2520	2.956	0.820							
	Teacher-Student Relationships	Intv	3767	3.298	0.923	-0.751	6282	.453	-0.018	0.024	-0.065	0.029
		Ctrl	2517	3.316	0.955							
	Valuing of School	Intv	3793	3.358	0.808	1.370	6327	.171	0.029	0.021	-0.012	0.069
	Teacher-Student Relationships	Ctrl	2536	3.330	0.815					0.024		
Teacher Climate	School Climate Educator version	Intv	261	3.312	0.602	1.337	516	.182	0.163	0.122	-0.077	0.403
		Ctrl	257	3.472	0.567							

Professional Learning	Intv	261	3.399	0.712	-1.167	515	.244	-0.075	0.064	-0.201	0.051
	Ctrl	256	3.474	0.748							
Feedback and Coahcing	Intv	260	2.925	0.803	-1.745	514	.082	-0.130	0.075	-0.277	0.016
	Ctrl	256	3.056	0.892							
School Leadership	Intv	261	3.479	0.788	-1.634	515	.103	-0.115	0.070	-0.253	0.023
	Ctrl	256	3.594	0.812							
Educating All Students	Intv	260	3.848	0.574	-0.987	514	.324	-0.054	0.055	-0.162	0.054
	Ctrl	256	3.903	0.670							
Teacher Self Reflection	Intv	259	3.827	0.599	-0.837	514	.403	-0.075	0.090	-0.252	0.102
	Ctrl	257	3.940	0.617							
School Safety - educator version	Intv	259	3.574	0.707	1.625	515	.105	0.114	0.070	-0.024	0.252
	Ctrl	258	3.460	0.880							

Note. Mean Difference = Mean (Intv) – Mean (ctrl), Intv = Intervention group; Ctrl = Control group.

Table 6

Independent t-tests of Outcome Variables at Post-Test Between the Treatment and Control Groups

Survey Post-	Outcomes at Test	Grp	*n*	*Mean*	*SD*	*t*	*df*	*p*	Mean Difference	Std. Error	95% CI of Mean Difference	
											Lower	Upper
Student Attribute	Grit	Intv	420	3.496	0.646	0.856	1054	.392	0.035	0.017	-0.031	0.036
		Ctrl	636	3.460	0.713							
	Growth Mindset	Intv	420	3.501	0.797	0.649	7085	.517	0.012	0.019	-0.025	0.049
		Ctrl	636	3.546	0.781							
	Learning Behaviors	Intv	420	4.069	0.611	0.417	7065	.677	0.006	0.015	-0.023	0.035
		Ctrl	636	3.999	0.630							
	Learning Strategies	Intv	420	3.468	0.768	1.257	7061	.209	0.023	0.018	-0.013	0.059
		Ctrl	636	3.445	0.771							
	Self-Efficacy	Intv	420	3.251	0.810	0.308	7095	.758	0.006	0.019	-0.032	0.044
		Ctrl	636	3.245	0.812							
	Self-Management	Intv	420	3.902	0.630	1.904	7089	.057	0.029	0.015	-0.001	0.058
		Ctrl	636	3.874	0.629							
	Social Awareness	Intv	420	3.580	0.646	0.225	7092	.822	0.004	0.016	-0.027	0.034
		Ctrl	636	3.576	0.664							

Student Climate	Engagement	Intv	400	2.911	0.839	1.199	6320	.230	0.026	0.022	-0.016	0.068
		Ctrl	658	2.885	0.849							
	Rigorous Expectations	Intv	3785	3.625	0.719	0.500	6316	.617	-0.009	0.018	-0.045	0.027
		Ctrl	2533	3.634	0.708							
	School Climate	Intv	3785	4.246	1.086	1.328	6318	.184	0.037	0.028	-0.018	0.093
		Ctrl	2535	4.209	1.111							
	School Safety	Intv	3771	3.461	0.773	1.456	5415	.145	0.046	0.020	0.006	0.086
		Ctrl	2521	3.415	0.819							
	Sense of Belonging	Intv	3770	2.987	0.824	1.455	6288	.146	0.031	0.021	-0.011	0.072
		Ctrl	2520	2.956	0.820							
	Teacher-Student Relationships	Intv	3767	3.298	0.923	-0.751	6282	.453	-0.018	0.024	-0.065	0.029
		Ctrl	2517	3.316	0.955							
	Valuing of School	Intv	3793	3.358	0.808	1.370	6327	.171	0.029	0.021	-0.012	0.069
	Teacher-Student Relationships	Ctrl	2536	3.330	0.815					0.024		
Teacher Climate	School Climate Educator version	Intv	261	3.312	0.602	1.337	516	.182	0.163	0.122	-0.077	0.403
		Ctrl	257	3.472	0.567							

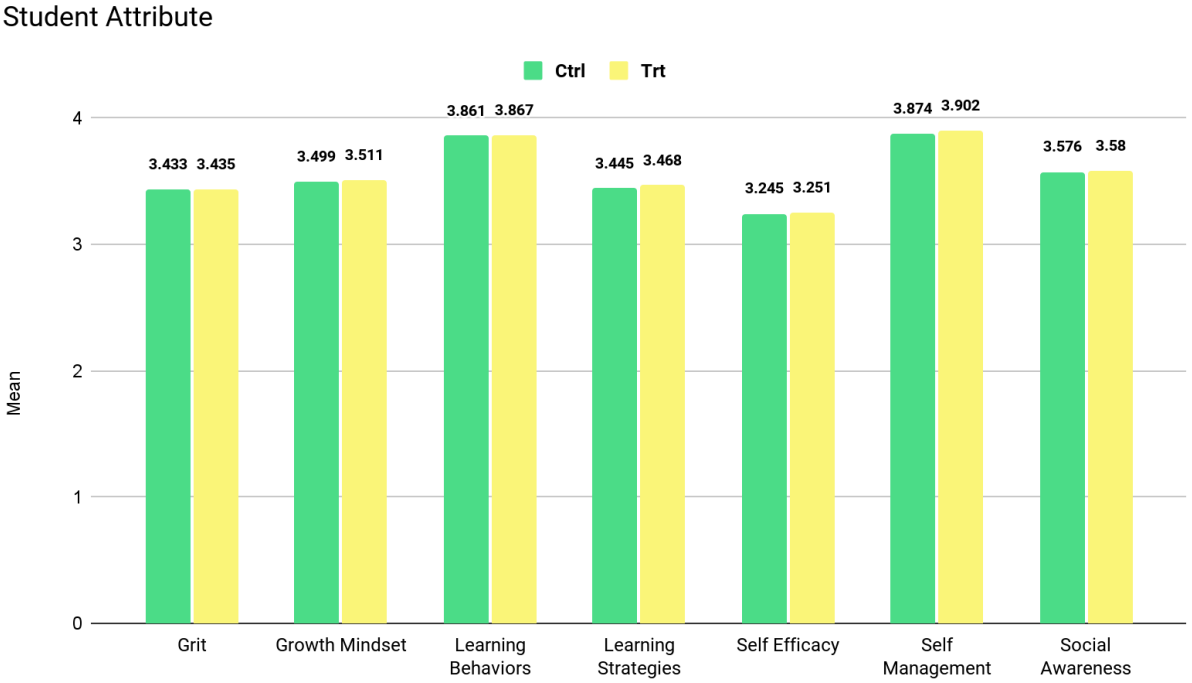
Professional Learning	Intv	261	3.399	0.712	-1.167	515	.244	-0.075	0.064	-0.201	0.051
	Ctrl	256	3.474	0.748							
Feedback and Coahcing	Intv	260	2.925	0.803	-1.745	514	.082	-0.130	0.075	-0.277	0.016
	Ctrl	256	3.056	0.892							
School Leadership	Intv	261	3.479	0.788	-1.634	515	.103	-0.115	0.070	-0.253	0.023
	Ctrl	256	3.594	0.812							
Educating All Students	Intv	260	3.848	0.574	-0.987	514	.324	-0.054	0.055	-0.162	0.054
	Ctrl	256	3.903	0.670							
Teacher Self Reflection	Intv	259	3.827	0.599	-0.837	514	.403	-0.075	0.090	-0.252	0.102
	Ctrl	257	3.940	0.617							
School Safety - educator version	Intv	259	3.574	0.707	1.625	515	.105	0.114	0.070	-0.024	0.252
	Ctrl	258	3.460	0.880							

Note. *p < .05. **p < .01. ***p < .001, indicates statistically significant. Mean Difference = Mean (trt) – Mean (ctrl), Intv = Intervention group; Ctrl = Control group.

Table 7

Chi-Square tests for ancillary outcomes.

Outcome	χ^2	*df*	*p*
frequency of serious misconduct	0.126	1	.723
Average Cumulative Absence rate - all codes	3.417	4	.491
Average Cumulative Absence rate - unexcused/truancy only	16.161	4	003



		Frequency	%	Frequency	%	Rate (%)
Studnt Attribute	Control	2536	40.07	636	60.2	74.92
	Treatment	3793	59.93	420	39.8	88.93
	Total	6329	100	1056	100	83.31
Student School Climate	Control	3270	45.98	658	62.1	79.88
	Treatment	3842	54.02	401	37.9	89.56
	Total	7112	100	1059	100	85.11

Teacher School Climate	Control	258	49.7	258	81.9	0	
	Treatment	261	50.3	57	18.1	78.16	
	Total	519	100	315	100	39.31	

Intv	420	4.021	0.583	1.103	1054	.270	0.039	0.035	-0.030	0.107
Intv	400	3.733	0.694	0.926	1056	.355	0.040	0.043	-0.045	0.126
Ctrl	658	3.6920	0.680							
Intv	400	3.982	0.996	0.750	1056	.453	0.046	0.062	-0.075	0.168
Ctrl	658	3.936	0.962							
Intv	400	4.125	0.669	3.942	1056	<.001***	0.169	0.043	0.085	0.253
Ctrl	658	3.956	0.680							
Intv	400	3.154	0.789	1.054	1056	.292	0.053	0.051	-0.046	0.153
Ctrl	658	3.100	0.804							

Figure 1. Student Attribute Survey at Pretest. Trt = Intervention group; Ctrl = Control group.

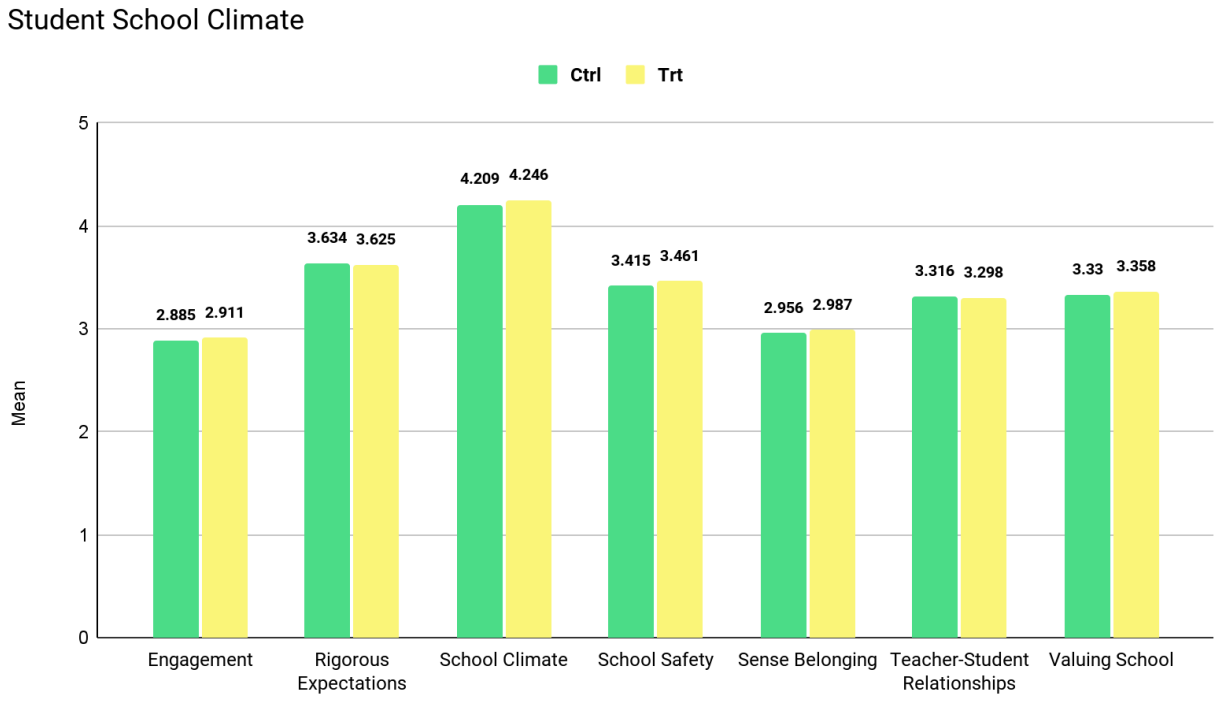


Figure 2. Student School Climate Survey at Pretest. Trt = Intervention group; Ctrl = Control group.

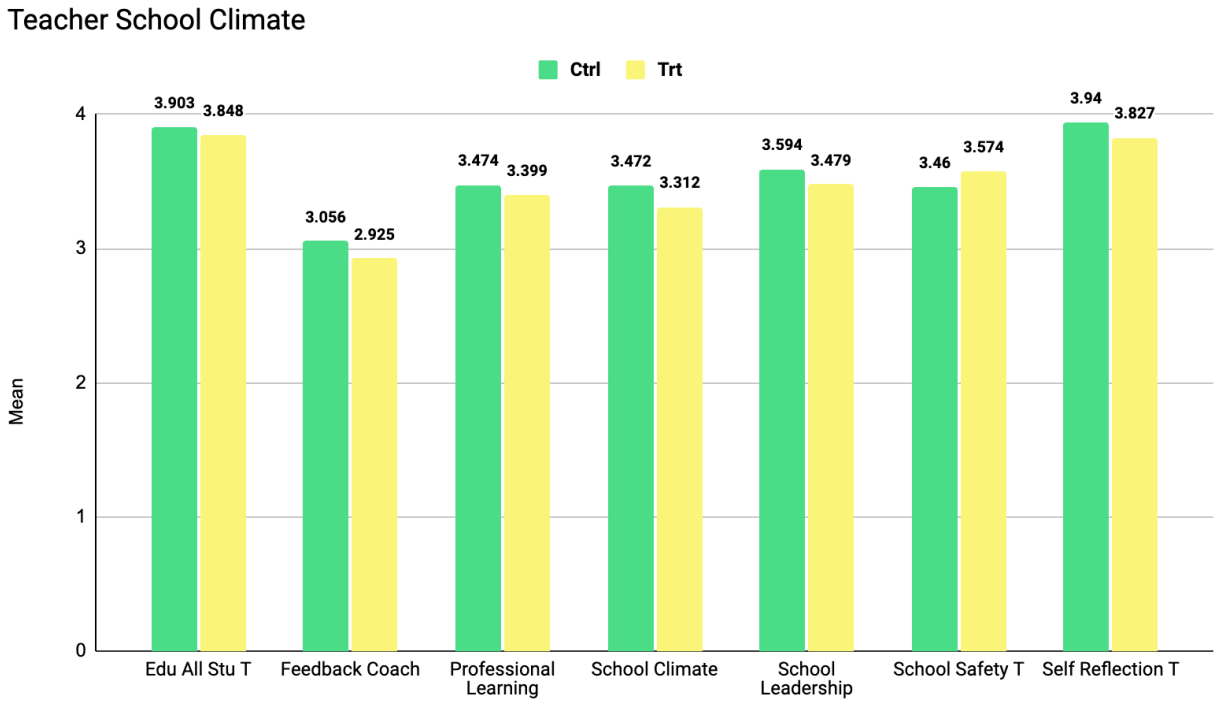
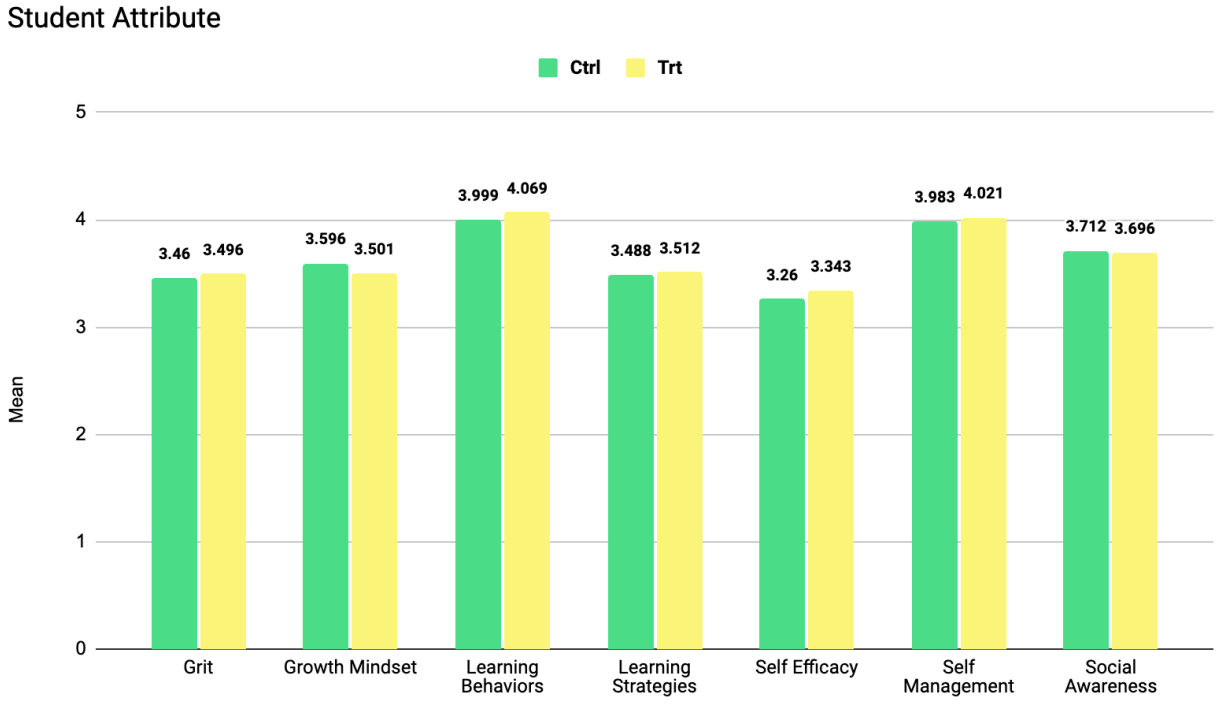


Figure 3. Teacher School Climate Survey at Pretest. Trt = Intervention group; Ctrl = Control group.



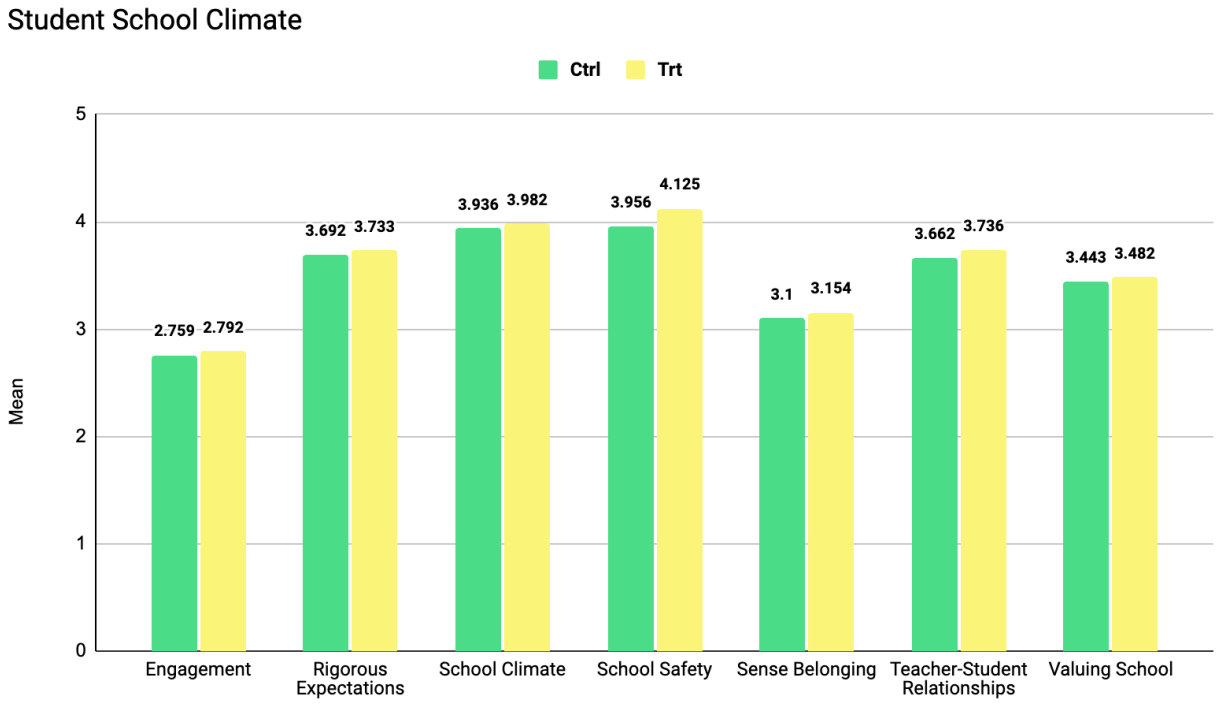


Figure 5. Student School Climate Survey at Posttest. Trt = Intervention group; Ctrl = Control group.

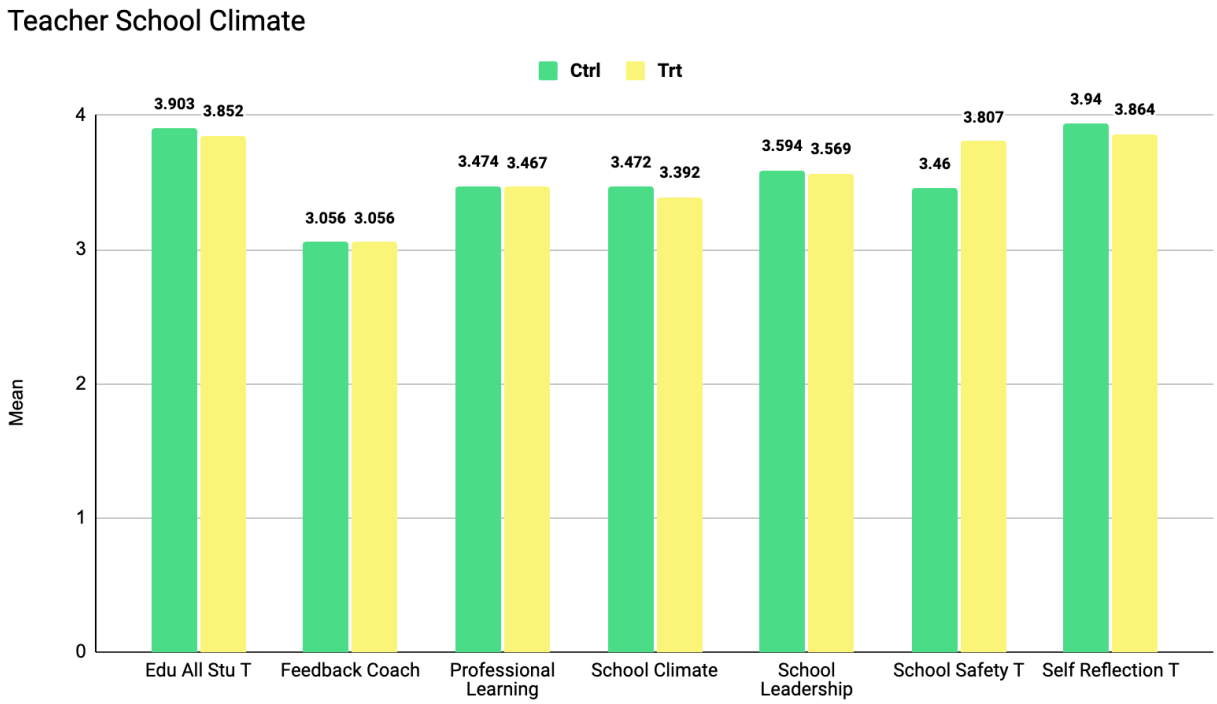


Figure 6. Teacher School Climate Survey at Posttest. Trt = Intervention group; Ctrl = Control group.