

Appendix 1

Survey Instrument

Our research team administered surveys at the beginning and end, following treatment, of each student's first year. The surveys were administered on paper and a member of our research team read aloud each question along with answer options while students completed their surveys. The survey also included demographic questions on students' age, race/ethnicity, and gender, which are not included in this appendix.

Art Consumption (Greene et al., 2018; Greene, Kisida, & Bowen, 2014) - students indicate whether they disagree a lot, disagree a little, do not agree or disagree, agree a little, or agree a lot with each statement.

Visual Arts

- Visiting art museums is fun.
- I plan to visit art museums when I am an adult.
- Art is interesting to me.
- I feel like I don't belong when I'm at an art museum.
- I feel comfortable talking about art.
- I would tell my friends that they should visit an art museum.
- How interested are you in visiting an art museum?
Students had different answer options for this item.
Not interested / Slightly interested / Somewhat interested / Interested / Very interested

Symphony

- Listening to orchestra music is interesting to me.
- I feel comfortable talking about orchestra music.
- I would tell my friends that they should hear an orchestra music concert.
- I plan to go to orchestra music performances when I am an adult.
- Orchestra music concerts are fun.
- How interested are you in going to an orchestra music performance?
Students had different answer options for this item.
Not interested / Slightly interested / Somewhat interested / Interested / Very interested

Theater

- Trips to see live theater are fun.
- Live theater is interesting to me.
- I feel comfortable talking about theater performances.
- I would tell my friends that they should see a live theater performance.
- I feel like I don't belong when I'm in a theater.
- I plan to see live theater performances when I am an adult.
- How interested are you in seeing live performances in a theater?

Students had different answer options for this item.

Not interested / Slightly interested / Somewhat interested / Interested / Very interested

Art Participation (Greene et al., 2018; Greene, Kisida, & Bowen, 2014) - students indicate whether they are not interested, slightly interested, somewhat interested, interested, or very interested to each statement.

Visual Arts

- How interested are you in making a work of art?
- How interested would you be in entering your work of art in a contest?
- How interested are you in taking an art class?
- I would be interested in joining an art club if my school had one.

Symphony

- If your school had an orchestra or band, how interested would you be in playing a musical instrument in it?
- How interested are you in taking music class?
- How interested are you in learning to play a musical instrument?
- I would be interested in joining an orchestra music club if my school had one.

Theater

- How interested are you in being in a theater performance?
- How interested are you in taking a drama class?
- If your school were having auditions for a play, how interested would you be in trying to get a role in that play?
- I would be interested in joining a drama club if my school had one.

Empathy- students indicate whether they disagree a lot, disagree a little, do not agree or disagree, agree a little, or agree a lot with each statement.

- It upsets me when another child is being shouted at.
- When I see someone suffering, I feel bad too.
- It makes me sad to see a child who can't find anyone to play with.

Fantasy Empathy Items (Davis, 1980).^{xv}

- After seeing a play or movie, I have felt as though I were one of the characters.
- When I watch a good movie, I can very easily put myself in the place of the leading character.
- When I am reading an interesting story or novel, I imagine how I would feel if the events in the story were happening to me.

Political Tolerance (Peterson, Campbell, & West, 2001)^{xvi}- students indicate whether they disagree a lot, disagree a little, do not agree or disagree, agree a little, or agree a lot with each statement.

- Some people have views that you oppose very strongly. Do you agree that these people should be allowed to come to your school and give a speech?
- Some people have views that you oppose very strongly. Do you agree that these people should be allowed to live in your neighborhood?

- Some people have views that you oppose very strongly. Do you agree that these people should be allowed to run for president?

Tolerance (Greene, et al., 2018; Greene, Kisida, & Bowen, 2014)- students indicate whether they disagree a lot, disagree a little, do not agree or disagree, agree a little, or agree a lot with each statement.

- I think people can have different opinions about the same thing.
- Women are equally able to do the same jobs that men can do.
- I am interested in learning about people different than me.

School Engagement^{xvii}- students indicate whether they disagree a lot, disagree a little, do not agree or disagree, agree a little, or agree a lot with each statement.

- Sometimes school is a waste of time.
- I feel proud being a part of this school.
- Getting good grades is important to me.
- School is boring.

Social Perspective Taking (Gehlbach, 2004; Gehlbach et al., 2008; Gehlbach, Brinkworth, & Wang, 2012; Greene et al., 2018)- students had the following answer options, almost never, once in a while, sometimes, often, or almost all the time

- How often do you attempt to understand your friends better by trying to figure out what they are thinking?
- How often do you try to think of more than one explanation for why someone else acted as they did?
- Overall, how often do you try to understand the point of view of other people?
- When you are angry at someone, how often do you try to "put yourself in his or her shoes"?
- How often do you try to figure out what motivates others to behave as they do?
- How often do you try to figure out what emotions people are feeling when you meet them for the first time?
- In general, how often do you try to understand how other people view the situation?

Appendix 2
Additional Tables

Table A1: Summary Statistics

	(1)	(2)	(3)	(4)	(5)
	Observations	Mean	SD	Minimum	Maximum
Treatment Variables					
Ever treatment	3,908	0.56	0.50	0	1
First treatment	3,908	0.31	0.46	0	1
Second Treatment	3,906	0.10	0.30	0	1
One-year post treatment	3,906	0.11	0.32	0	1
Two-years post treatment	3,908	0.03	0.16	0	1
Demographics					
Female	3,896	0.53	0.50	0	1
SWD	3,782	0.14	0.34	0	1
Black or African American	3,889	0.93	0.26	0	1
Pre- and Post- Treatment Measures					
Pre-combined test score	3,425	0	1	-2.78	4.15
Post-combined test score	3,473	0	1	-2.50	3.56
Pre-course grades	3,461	0	1	-5.06	2.71
Post-course grades	3,547	0	1	-5.88	3.06
Pre-proportion days absent	3,141	0.03	0.03	0	0.13
Post-proportion days absent	3,141	0.04	0.03	0	0.15
Pre-number of infractions	3890	0.14	0.61	0	7
Post-number of infractions	3890	0.35	0.98	0	7
Pre-art consumption	3,463	0	1	-3.47	1.62
Post-art consumption	2,061	0	1	-2.99	1.88
Pre-art participation	3,463	0	1	-3.31	1.92
Post-art participation	2,061	0	1	-2.98	1.97
Pre "different opinions"	3,418	0	1	-3.61	0.55
Post "different opinions"	2,049	0	1	-3.73	0.56
Pre- SPT	3,436	0	1	-2.79	2.00
Post- SPT	2,050	0	1	-2.61	2.11
Pre-empathy	3,462	0	1	-3.47	1.02
Post-empathy	2,060	0	1	-3.20	1.11
Pre-careless answers	3,463	0	1	-2.40	2.96
Post-careless answers	2,060	0	1	-2.61	3.19
Pre-item non-response	3,451	0	1	-1.60	16.02
Post-item non-response	2,054	0	1	-3.10	19.01

Notes: Table includes summary statistics for dependent and independent variables. All student observations over the three years are included. There is a total of 2,197 individual students with 968 as control and 1,229 as treatment. Most outcome variables are in standard deviations, except for the number of infractions and the proportion of days absent from school. All test scores are standardized Georgia Millstone end-of-grade exams and are standardized within grade level by year. We removed a small number of outliers in the number of infractions and the proportion of days absent, accounting for less than 0.5 percent of the sample in the infraction outcome analysis and 5 percent of the sample in the attendance outcome analysis.

Table A2: Treatment Effect by Cohort on Test scores, Course Grade, Attendance, and Infractions

	1st Treatment (1)	2nd Treatment (2)	1 Yr Post Treatment (3)	2 Yrs Post Treatment (4)	# of Observations (6)	# of Students (7)
<i>Panel A: Cohort 1</i>						
Combined Test Scores	0.113* (0.060)	0.201*** (0.076)	0.152*** (0.038)	0.206*** (0.051)	1082	434
Course Grades	-0.163 (0.121)	0.090 (0.160)	0.159 (0.097)	0.252** (0.122)	1108	434
Proportion Absent	-0.000 (0.003)	-0.001 (0.005)	-0.005 (0.003)	0.005 (0.005)	1056	429
# Behavioral Infractions	0.002 (0.102)	0.052 (0.152)	-0.291** (0.117)	0.019 (0.190)	1229	457
<i>Panel B: Cohort 2</i>						
Combined Test Scores	-0.040 (0.061)	-0.039 (0.069)	-0.135* (0.070)	- -	1142	610
Course Grades	-0.104 (0.099)	-0.172 (0.142)	-0.204 (0.157)	- -	1155	616
Proportion Absent	0.002 (0.002)	0.007** (0.003)	-0.007* (0.004)	- -	1152	619
# Behavioral Infractions	0.148* (0.089)	0.360*** (0.112)	-0.040 (0.132)	- -	1198	638
<i>Panel C: Cohort 3</i>						
Combined Test Scores	0.016 (0.050)	- -	- -	- -	620	620
Course Grades	0.282** (0.111)	- -	- -	- -	624	624
Proportion Absent	-0.007*** (0.002)	- -	- -	- -	611	611
# Behavioral Infractions	-0.104 (0.126)	- -	- -	- -	644	644

Notes: *** p<0.01, ** p<0.05, * p<0.1. Each estimated treatment effect comes from separate regressions. Standard errors clustered at the classroom level are in parentheses. All models include school, cohort, and year fixed effects, student random effects, along with controls for students' gender, SWD status, baseline combined standardized test scores, pre-treatment desire to consume arts, the grade the student was randomized in (fourth or fifth grade), and the baseline measure of the given outcome. Models for cohort 1 and 2 also include student random effects. Combined test score is a standardized score of a student's Georgia Milestone ELA and math exam. Test scores were standardized within grade.

ⁱ The fifteen elementary schools were selected by The Woodruff Arts Center to participate in the study and are not necessarily representative of the larger school district.

ⁱⁱ Details of the field trips were provided by the Woodruff Arts Center as well as our research teams' observations of the field trips. For more information about The Woodruff and the three arts partners see <https://www.woodruffcenter.org/>.

ⁱⁱⁱ This is not a comprehensive list of field trips that students in the control group attended, as the specific field trip varied by school and grade.

^{iv} Cohort one treatment students who are in sixth grade in year three, received treatment in both fourth and fifth grade. As such, in year three, these students are one-year post a second treatment. We do not estimate the effect of one-year post one treatment and one-year post two treatments because there is a limited number of students who fall into these subcategories. We only estimate the effect of one-year post treatment regardless of if students received one or two dosages of the treatment in previous years.

^v The survey included separate constructs for art participation and consumption for each art institution, an art museum, theater, and symphony. We combine the three subcategories into one overall art consumption scale and one art participation scale. Refer to Appendix A for the complete constructs.

^{vi} The Georgia Milestones end-of-grade exams are used as a significant part of the state's school accountability program and have both norm-referenced and criterion-referenced items. Students in grades three through eight are required to take both the math and ELA exam. Students in fifth through eighth grade also take a science and social studies exam. As not all student in our sample take the science and social studies exams, we only focus on math and ELA scores. We standardize all exam scores within grade level by year.

^{vii} We do not estimate any treatment on the treated effect as we do not have data on which students actually attended the various field trips.

^{viii} We use the term baseline measures when referring to administrative data, but when referring to survey data, we use the term pre-treatment measures as pre-treatment surveys were collected after randomization.

^{ix} Refer to appendix B, Table 1B for summary statistics on all variables.

^x Based on our conversations with teachers, teachers who were assigned to treatment made greater effort to collect consent forms than control teachers. The additional effort from treatment teachers could create dissimilar treatment and control groups. We believe this would create bias against our treatment group, making it more difficult to detect an effect. Because the treatment group would include students who were less motivated or less conscientious than the control group. While the treatment and control groups might differ on unobservable characteristics, we observe few significant differences at baseline, despite the differential consent.

^{xi} We combine students' math and ELA test scores for an overall measure of performance on the Georgia Milestones end-of-grade exams. The treatment effect remains positive and significant for both math and ELA exam scores when we run separate regressions for each subject.

^{xii} We excluded just under five percent of our sample due to a number of outliers in the attendance data for whom we observed very high rates of absences, with some students missing 20 to 75 percent of the school days. We originally excluded students from our sample whose absent rate was two standard deviations above the mean. On further robustness checks, the results are not robust to different exclusion rules.

^{xiii} The average number of infractions for elementary students in our sample is 0.264 with a standard deviation of 0.812 compared to middle school students with 0.608 average infractions and a standard deviation of 1.241.

^{xiv} Refer to Appendix B, Table 2B for the estimated treatment effect on all administrative outcomes by cohort.

^{xv} The fantasy empathy subscale was added in the second year of the study.

^{xvi} The political tolerance scale was added in the second year of the study.

^{xvii} In the first year of the study, only the item “School is boring.” was included on the survey. In the second year of the study we added the remaining items.