Investigating the Causal Effects of Arts Education

Daniel H. Bowen
Brian Kisida

Abstract

There is a paucity of research on the causal relationship between arts learning and educational outcomes. Investigating these relationships has become imperative as policymakers increasingly prioritize empirical evidence of educational impacts, which often leads to curriculum narrowing that favors traditionally-tested subjects. Employing a randomized controlled trial with 42 elementary and middle schools in Houston, Texas, we find that randomly assigning arts educational opportunities reduces disciplinary infractions, improves writing achievement, and increases students' emotional empathy. Students in elementary schools, which were the primary focus of the program, also experience increases in school engagement, college aspirations, and cognitive empathy. As the first large-scale randomized control trial of arts learning in an authentic school setting, these findings provide strong evidence that the arts can produce meaningful impacts on students' academic outcomes and social-emotional development. Education policymakers should consider these benefits when assessing the role of the arts in schools. © *2022 by the Association for Public Policy Analysis and Management.*

INTRODUCTION

The arts maintain a precarious position in public education. After a steady increase throughout the middle of the 20th century, arts education in the United States has been in steady decline since the 1980s (Rabkin & Hedberg, 2011). School accountability systems that emphasize achievement on standardized tests have precipitated this decline (Center on Education Policy, 2008; Gadsden, 2008). School systems have responded to accountability pressure by prioritizing instruction aligned with high-stakes assessment measures, leading to a reduction in scarce resources allocated to untested subjects (Bassok et al., 2016; Dee et al., 2013; West, 2007). In a 2012 national survey, roughly half of public school teachers reported art and music receiving less instructional time and resources over the previous decade. By contrast, few teachers indicated reductions in math (10 percent) and English language arts (12 percent) (Farkas Duffett Research Group, 2012). When asked to indicate the main driver of reductions, 36 percent cited test pressure, 28 percent cited budget cuts, and 29 percent cited both equally.

These reductions have not been equitable across students and schools. Test-based accountability pressures and budget constraints have been more severe at schools serving historically-underserved students, specifically those serving higher proportions of lower-income, Black, Hispanic, and immigrant students (Balfanz et al.,

Journal of Policy Analysis and Management, Vol. 0, No. 0, 1–24 (2022)

Published by Wiley Periodicals, Inc. View this article online at wileyonlinelibrary.com/journal/pam
Additional supporting information may be found online in the Supporting Information section at the end of the article.
DOI:10.1002/pam.22449

2007; Government Accountability Office, 2009). From the 1980s through the first decade of the 20th century, Black and Hispanic students experienced 25 and 19 percentage point declines in arts education participation, respectively. Conversely, White student participation decreased by one percentage point (Rabkin & Hedberg, 2011). Schools with higher shares of free and reduced-price lunch (FRL) eligibility also have consistently less access to the arts in terms of instructional time, arts specialists, course offerings, and dedicated facilities (Government Accounting Office, 2009; Parsad & Spiegelman, 2012). It has become increasingly common for under-resourced schools to offer little to no arts instruction (Yee, 2014).

Recent trends may provide new opportunities for the preservation and restoration of arts education. The arts received equal footing alongside math and reading in the most recent (2015) reauthorization of the Elementary and Secondary Education Act's definition of a well-rounded education, and new requirements for broader accountability measures have prompted states to include indicators such as school engagement, disciplinary infractions, attendance, social-emotional learning, and school climate when assessing school quality (Batel, 2017). This corresponds to a growing research base demonstrating the effects of these broader indicators on long-run student outcomes, such as improving high-school completion and college attendance (Duckworth et al., 2007; Jackson, 2018; Jackson et al., 2020; Kraft, 2019; Soland et al., 2013). This trend bodes well for the arts. Though there are numerous dubious claims that the arts improve traditionally assessed educational outcomes, such as math and reading achievement, there is little causal evidence (Education Commission of the States, 2019; Ludwig et al., 2017; Wan et al., 2018; Winner et al., 2013). Theory and emerging research, however, suggest arts education is more likely to affect behavioral and social-emotional learning outcomes. Rigorously identifying this causal link would strengthen the role of arts education in a policy environment increasingly focused on a broader set of quantifiable benefits.

Current strategies to retain arts education in under-resourced schools have taken numerous forms, including integrating the arts into other subjects, supplementing resources with philanthropic funding, and forming partnerships with cultural and community organizations. According to the U.S. Department of Education, 42 percent of U.S. public schools partner or collaborate with cultural or community organizations, 31 percent with individual artists, 29 percent with museums, and 26 percent with performing arts centers (Parsad & Spiegelman, 2012). In many cases, schools and districts form coalitions with multiple arts and community partners to address declines in arts resources (Bowen & Kisida, 2017; Perille, 2016). These initiatives vary in their scope and employ a broad range of strategies, including increasing public school arts education funding and course availability through public advocacy, outreach efforts to build community support, awarding grants to schools and classroom arts educators, and cultivating partnerships with cultural organizations that provide educational services (Gibson, 2016; Perille, 2016). Examples of these formal coordinated efforts can be found in many large school urban districts, including Boston, Chicago, Dallas, Los Angeles, New Orleans, and Seattle.

In this study, we investigate the causal effects of one such multi-sector collaborative effort that substantially increased arts learning opportunities in under-resourced schools, Houston's Arts Access Initiative (AAI). The AAI coalition included the Houston Independent School District (HISD), a broad set of more than 50 cultural institutions and arts education organizations, representatives from the mayor's office, local philanthropies, researchers, and a "backbone" organization that facilitated the collaborative effort. Employing a randomized controlled trial (RCT) of 42 participating elementary and middle public schools, we investigate the AAI's effects on a host of policy-relevant educational outcomes: discipline, attendance, standardized test scores (reading, math, writing, and science), college aspirations, and indicators of social-emotional development: school engagement, empathy, and tolerance.

This is the first large-scale experimental investigation of an increasingly common arts education intervention that is being implemented in authentic school settings across the United States.

We find that arts learning reduces the proportion of students receiving disciplinary infractions by 3.6 percentage points, increases standardized writing achievement scores by 0.13 of a standard deviation, and increases students' emotional empathy by 0.07 of a standard deviation. We do not find overall significant effects on student attendance, reading, math, or science achievement scores, or other survey outcomes. Students in elementary schools, which comprised 36 of 42 schools in our sample and were the primary focus of the program, also experience significant increases in school engagement, college aspirations, and cognitive empathy. Results are particularly pronounced for English language-learning students in elementary schools. These findings provide strong evidence that the arts can produce a variety of meaningful impacts on students' educational outcomes.

BACKGROUND

The benefits of arts education are rich in theory and testimony, but little rigorous evidence supports most claims. Numerous studies find positive relationships between arts participation and commonly measured educational outcomes, such as standardized test scores, graduation rates, and college enrollment (Catterall et al., 2012; Elpus, 2013). Such studies, however, tend to generate skepticism on two fronts. First, they generally lack rigorous research designs and do little more than establish correlations between arts course-taking and supposed benefits (Winner & Cooper, 2000). Second, they rely on outcomes, like test scores in other subjects, with little theoretical relationship to arts education. To the extent such studies are believed, they can generate unrealistic expectations that arts education is unlikely to meet (Winner et al., 2013).

More broadly, the nascent field of arts education research has yet to settle on a well-defined theory of change that connects arts education experiences to anticipated and measurable outcomes (McCarthy et al., 2004). This is due in part to the multifaceted nature of arts learning across different arts disciplines, as well as the difficulty in assessing outcomes not commonly found in administrative data.

A recent report from the American Academy of Arts and Sciences' (2021) two-year Commission on the Arts took stock of the many theories and claims surrounding arts education and identified seven areas of educational benefits that are supported by theory and research. First, there is the primary claim that learning about the arts is good for its own sake. The arts are a fundamental mode of human expression and, like other subjects, are a "way of knowing and understanding the world" (American Academy of Arts and Sciences, 2021, p. 11; Winner et al., 2013). Moreover, familiarity with the arts helps students acquire cultural capital, which has longer-term social and economic benefits (Bourdieu, 1977; Roksa & Potter 2011; Roscigno & Ainsworth-Darnell, 1999).

The other six thematic areas identified in the Commission on the Arts (2021) report are what some would consider intrinsic arts benefits (McCarthy et al., 2004). That is, though not direct measures of arts content and skills, they are intrinsically linked to the process of learning about and through the arts. The report asserts that arts education broadens students' understanding of other cultures and history; supports social-emotional development, such as facilitating self-expression and emotional growth; generates prosocial/interpersonal skills, such as empathy and acceptance of others; improves school engagement, connectedness, and culture; provides career exploration and skills, including hands-on skills as well as critical thinking skills and creativity; and strengthens community and civic engagement. Notably

absent are claims that the arts directly affect student achievement in other tested subjects, though it remains plausible that such "transfer" effects might occur as an indirect result of arts' mediating influence on school engagement and enjoyment. Recent research finds social-emotional development explains a considerable amount of schools' impacts on test scores (Jackson et al., 2020). Arts education may also directly affect achievement in other tested subjects if certain skills, such as writing skills derived from literary arts courses, transfer to improvements in general academic abilities.

Social-emotional and school engagement outcomes related to the arts have particular policy relevance given the increased attention they have received in recent years and emerging research demonstrating their significance (Jackson, 2018; Jackson et al., 2020; Kraft, 2019). The arts provide unique opportunities for self-discovery, social engagement with peers, and exposure to a broader world of ideas and topics that can spark student interest in learning (Dewey, 1919; Konrath & Kisida, 2021; Mehta, 2017). Research examining schools' effects on such skills and mindsets find that they strongly predict later-life outcomes, often more than test scores (Jackson, 2018; Jackson et al., 2020). Schools that improve students' social-emotional development—as measured by survey items capturing interpersonal skills, school connectedness, and academic engagement—increase the likelihood that students are on-track; have fewer absences, disciplinary infractions, and school-based arrests; and increase graduation rates, four-year college-going, and persistence in college (Jackson 2018, Jackson et al., 2020).

Thus far, causal studies of arts experiences' impacts on social-emotional development have been investigated primarily through experimental studies of arts-based field trips. Students randomly assigned to visit art museums or live theater performances have demonstrated increased empathy and tolerance for others (Greene et al., 2014; Greene et al., 2018; Kisida et al., 2020). The arts-based field trip literature also finds some evidence of improved behavioral outcomes that proxy for school engagement, such as improved attendance and disciplinary outcomes (Erickson et al., 2022; Lacoe et al., 2020). There is also suggestive but limited evidence that these arts experiences may have transfer effects on more direct academic measures. For example, students randomly assigned to participate in an art museum's educational program subsequently displayed stronger critical thinking skills when composing essays about works of art (Bowen et al., 2014; Kisida et al., 2016). In another study, students randomly assigned to participate in a program that integrated historical content with live theater demonstrated increased interest in learning about history and improved content knowledge aligned with state history standards (Kisida et al., 2020). To date, however, there is a lack of rigorous research on the benefits of arts education in authentic school settings.

There are also lingering policy-relevant questions regarding who most benefits from school-sponsored arts education activities. Race/ethnicity and socioeconomic status are strong predictors of access to arts and cultural experiences outside of school, with historically-underserved populations receiving substantially less exposure through family-facilitated experiences (American Academy of Arts and Sciences, 2021; Greene et al., 2014; Meyer et al., 2004; Redford et al., 2018). Consequently, students from underserved communities are more dependent on schools to provide arts access, and may respond more to interventions that fill that need. Limited research confirms this expectation, as historically-underserved students have demonstrated more-pronounced positive relationships from school-sponsored arts exposure (Catterall et al., 2012; Greene et al., 2014; Kinney & Forsythe, 2005; Podlozny, 2000; Thomas et al., 2015). In a similar vein, a growing body of research finds that arts integration (i.e., integrating arts content and techniques into other subjects) has particular benefits for English-language learners and students who perform lower on standardized tests (Hardiman et al., 2019; Peppler et al., 2014).

We advance this literature in several important ways. First, we conduct the first randomized study of a broad-based initiative that leverages community partners to deliver a substantial increase in arts education experiences in an authentic school setting. Similar interventions are occurring in many school districts throughout the country and evidence of their effectiveness is particularly policy-relevant. Second, we examine arts education effects on traditional educational outcomes derived from administrative data (attendance, discipline, and test scores), as well as survey items intended to capture the effects of arts education on a range of social-emotional learning outcomes, specifically school engagement, emotional and cognitive empathy, tolerance, and interest in pursuing higher education. Third, we conduct exploratory analyses on a broad range of policy-relevant subgroups, including identified race/ethnicity, poverty status, grade level, and English language learner status, to better understand the populations most affected through an arts-learning intervention.

School-Community Partnerships

Broad-based coalitions with multiple arts and community partners are an increasingly popular method that school districts engage in to address declines in arts resources (Bowen & Kisida, 2017; Perille, 2016). These initiatives employ a broad range of strategies, including working with policymakers to increase arts education funding and course availability, coordinating efforts to build community and philanthropic support through outreach, directly awarding grants to schools and classroom arts instructors, and cultivating partnerships with arts and cultural organizations to provide arts education experiences (Gibson, 2016; Perille, 2016). Examples of these coordinated efforts can be found in many large urban districts, including Boston (Arts Expansion Initiative), Chicago (Creative Schools Initiative), Dallas (Learning Partners), Los Angeles (Arts Education Collective), New Orleans (KID smART), and Seattle (Creative Advantage). Educational offerings provided by local arts organizations are a key ingredient in these initiatives. Common offerings include in-school teaching-artist residencies, workshops for students and teachers, professional artist performances, after-school programs, and field trips to arts institutions.

Houston's Arts Access Initiative

Launched in 2013, Houston's Arts Access Initiative (AAI) exemplified one such multisector collaborative effort. The AAI included the Houston Independent School District (HISD), a broad set of more than 50 cultural institutions and arts education organizations, representatives from the mayor's office, local philanthropies, researchers, and a "backbone" organization that facilitated the collaborative effort. The Initiative's focus was increasing arts access for students, with an emphasis on closing opportunity gaps between under-resourced and wealthier schools. A critical early step in the AAI was the collection of school-level arts educational resource data throughout HISD. The AAI coalition then assessed arts educational resources for each of HISD's elementary and middle schools in terms of the number of certified arts specialists, before- and after-school arts programs, and the number of partnerships with arts organizations over the prior school year. This inventory revealed that 29 percent of the 209 elementary and middle schools had no full-time arts specialist; 30 percent did not provide any arts programming outside of regular school hours; and 39 percent had either one or no arts partnerships with community arts organizations.

The findings from the school inventory motivated AAI stakeholders to prioritize recruitment of HISD elementary and middle schools with the lowest levels of arts resources to participate in the first two years of implementation. School participation in the Initiative was voluntary, and to be considered principals had to commit to its mission, engage in strategic arts planning with the AAI director, designate an arts liaison to coordinate AAI-related efforts, participate in teacher and principal arts-integration professional development, and attend AAI peer-network mentoring sessions.

Participating principals were required to commit between $1 to $10 per student from their schools' budgets to fund the main component of the AAI: arts educational experiences delivered through partnerships with arts organizations that provided teaching-artist residencies, on-campus and off-campus professional artist workshops and performances, after-school programs, and field trips to cultural arts organizations (see Appendix Table A1 for a complete list of partnering arts organizations by type and modality).[1] A private local philanthropic institution provided a 1:1 match for each school's financial commitment.

AAI leaders designed a gradual rollout strategy, with capacity for seven schools to participate in the first year and an additional 14 schools in the second year, after which the program became available to all eligible schools that applied. AAI administrators decided to allocate 18 of the 21 initial slots to elementary schools, as middle schools were more likely to have at least one certified arts specialist and had more arts resources and opportunities. After consulting with us, AAI facilitators agreed to randomly assign participation among the neediest schools in the first two years.

Demand for AAI participation in the first two years exceeded supply. Sixty eligible schools applied to participate in the AAI during this rollout phase: 32 in the first year, and an additional 28 in the second year. With an expected final sample size of 42 schools, we stratified randomization to assign treatment status to improve precision (Gerber & Green, 2012). Twenty-two of the 32 first-year applicants were paired based on decision rules that first grouped applicants based on school grade level (elementary or middle), then student demographics (percent of students FRL-eligible, Hispanic, Black, and White), preexisting arts resources (indicators for whether schools had certified arts teachers and number of arts partnerships), and finally baseline academic achievement (percentages of students "proficient" on standardized math and reading assessments). Ten first-year-applicant schools were not paired due to being too dissimilar from other applicants on these matching criteria. Seven of the 11 pairs were then randomly selected to participate in the first year of the evaluation and randomly assigned to treatment and control status within each matched pair. The 18 schools not assigned to the treatment or control group were encouraged to apply again next year. For AAI's second year of implementation, 28 applicants were paired and randomized using the same process from the previous year.

After random assignment, the AAI director and staff worked with the treatment schools' principals to formulate goals for the upcoming year and help guide arts program selections that would align with their educational objectives. As part of this process, the director and staff encouraged principals to budget for a diversity of programs such that all major arts disciplines were included: dance, music, theater, and visual arts. AAI personnel monitored and documented schools' arts expenditures to ensure that they fulfilled their financial commitments. The director and staff also kept and distributed the philanthropic institution's matched funds. Including these

[1] All appendices are available at the end of this article as it appears in JPAM online. Go to the publisher's website and use the search engine to locate the article at http://onlinelibrary.wiley.com.

matched funds, AAI schools had an average annual budget of $14.67 per student to facilitate partnerships with arts organizations. The average AAI school contribution was a mere 0.08 percent of the $9,663 of HISD's average per-pupil funding during the study period. It is also worth noting, however, that it was common for arts organizations to provide fully subsidized or discounted programming to AAI treatment schools that was supported through other fundraising efforts.

According to the baseline arts inventory, treatment and control schools were statistically equivalent in terms of their average number of school-community partnerships in the year prior to application. AAI treatment schools had 2.76 arts partnerships compared to 2.80 for the control group. Post-treatment, AAI schools averaged a net increase of 5.03 additional partnerships relative to the control schools over the same period. From these partnerships, AAI treatment schools participated in 9.86 additional arts educational experiences over the course of a school year.

The arts educational experiences students were exposed to were diverse and varied across participating schools. Fifty-four percent of these experiences were primarily theatre-based, 12 percent dance, 18 percent music, and 16 percent visual arts; 33 percent of these experiences were provided through teaching-artist residencies, 31 percent were in-school professional artist performances and workshops, 27 percent were field trip experiences, and 9 percent were programs provided outside of regular school hours. Teaching-artist residencies, the most commonly selected experience by participating schools, most closely approximated the role of a full-time arts specialist. Teaching-artists from participating arts organizations established semester- or year-long relationships with schools and provided arts instructional content to entire grades on a weekly or semi-weekly basis. Residencies were especially popular in elementary schools, as they were less likely to have full-time arts specialists prior to the initiative. In-school workshops or performances constituted the second most chosen experiences. These experiences could range from an hour-long performance or workshop to multi-day events. Off-campus performances and field trip experiences ranged from traditional one-off visits to major art museums or performance centers—such as the Museum of Fine Arts Houston, the Houston Ballet, Alley Theatre, and the Houston Symphony—or visits to medium- and smaller-sized community arts centers. Many of these activities also went beyond mere tours and included workshops and other learning activities.

The diversity of the treatment group's programming was as vibrant and multifaceted as the nonprofit arts and cultural sector. That said, there are three common characteristics worth noting. First, the educational mission of participating arts organizations was not happenstance. Rather, the vast majority have well-articulated educational philosophies and goals and had been providing educational services in some capacity since well before the creation of the AAI. Many were making deliberate efforts to align with state educational standards. For example, one chamber music organization stated, "The mission of the program is to establish creative learning environments that support the Texas Essential Knowledge Skills (TEKS)" (Da Camera, 2022). Such statements are typical. Besides claims of alignment with state learning standards, many organizations also integrated content from tested subjects. For example, one literary arts organization offered "Proven, TEKS-aligned creative writing workshops for K-12 students taught by professional writers" that "align with ELA, Social Studies and CTE core content" (Writers in the Schools, 2022). Another organization's workshops integrated science with classical music "to introduce the science of Galileo, Sir Isaac Newton, and Einstein," or civics by "hearing the favorite tunes of Ben Franklin…while learning about democracy and the people who helped create our nation" (Mercury Chamber Orchestra, 2022). Another program, Literacy through Photography (Fotofest, 2022), self-described as a "full-scale photography, writing, and literacy program."

In addition to traditional curricular goals, many arts organizations deliberately focused on social-emotional development. For example, one organization described themselves as "A forward-thinking music education and social-emotional learning initiative" (Nameless Sound, 2022). Another provider noted that "the artistic process enables students to develop the social skills and emotional intelligence attributed to high academic performance" (Multicultural Education and Counseling through the Arts, 2022).

Second, the programs are remarkably culturally diverse. While many participating arts organizations provided broad and diverse coverage spanning ancient, modern, and contemporary art forms that one would expect from major art museums and performance centers, many of the small- to medium-sized providers focused on specific cultural heritages. These providers include arts organizations that specialized in African dance and drumming, Asian dance, Aztec dance, Brazilian music and dance, Chinese art, Mexican *folklórico*, hip-hop music and dance, and Hispanic literature.

Finally, the vast majority of arts organizations had stated missions and experience focusing on underserved student populations, noting, for example, that their programs "provide social benefits by serving as an equalizer against socioeconomic barriers that may hinder a child's access to enriching arts activities" (Multicultural Education and Counseling through the Arts, 2022).

While the many diverse approaches and experiences are difficult to summarize, these examples provide a sense of the deliberate educational goals to enrich core content and social-emotional development, the vast array of cultural representation, and how programs were tailored for underserved students.

DATA

We obtained administrative data from HISD records through the Houston Education Research Consortium's database, which contains student-level school attendance and enrollment records; State of Texas Assessments of Academic Readiness (STAAR) standardized test scores; disciplinary records; and a rich set of student-level demographics including grade level, gender, race/ethnicity, FRL eligibility status, poverty status, and indicators for whether students received special education and English language learning (ELL) services. While all 3rd through 8th grade students take the STAAR reading and math assessments, only 4th and 7th grade students take the STAAR writing assessment, and only 5th and 8th grade students take the science assessment.

We collected original survey data in the second year of AAI's implementation—the year with the largest sample of schools during the evaluation. Using unique student identification codes, we linked survey data to student-level administrative records. HISD restricted survey participation to students enrolled in STAAR-tested grades (3rd through 8th) and requested that we exclude students who received special accommodations when taking state assessments. HISD also prohibited us from surveying students who transferred out of AAI evaluation schools, whether due to having graded out after the first year of implementation or transferring to a non-evaluation school. Schools' testing coordinators administered our baseline survey at the beginning of the fall semester (late September through early October) and the outcome survey at the end of the school year (late April through May). Students' responses on the spring survey provide outcome measures for both AAI evaluation cohorts. We control for fall baseline survey responses for cohort two students as a robustness check of our main findings.

Our survey items were developed in collaboration with AAI stakeholders and relied on arts education literature and survey instruments used in previous research.

These survey items were intended to capture levels of college aspirations and indicators of social-emotional learning: school engagement, emotional empathy, cognitive empathy, and tolerance. The survey items were rated on a four-category Likert scale where response choices ranged from "strongly disagree" to "strongly agree."

To reduce the number of outcomes we examine, we created constructs from the school engagement and cognitive empathy items to serve as dependent variables through exploratory factor analysis. The school engagement construct captures how students rate their agreement with statements about their interest and enjoyment of school, which includes the following five items (Cronbach's alpha = 0.72): "School work is interesting"; "This school is a happy place for me to be"; "School work makes me think about things in new ways"; "This school offers lots of different types of programs, classes, and activities to keep me interested in school"; and "School work is not very enjoyable" (reverse coded).

The empathy survey items were intended to capture the two most common components found in the empathy literature: cognitive empathy, which is understanding and learning from another's perspective, i.e., empathic cognition; and emotional empathy, which captures the emotional reaction to sharing in another's feelings, i.e., empathic concern (Cotton, 2001; Garton & Gringart, 2005; Martingano & Konrath, 2022). Our cognitive empathy construct measures students' competence at relating to others through arts experiences in terms of social connections to peers, perspective taking, and historical understanding. It contains the following three items (Cronbach's alpha = 0.65): "I can learn about my classmates by listening to them talk about works of art"; "Works of art help me imagine what life is like for someone else"; and "Works of art…help me understand what life was like in another time or place."

Cronbach's alphas for emotional empathy and tolerance items, however, had reliability indicators below 0.6. As a result, we do not report results for these constructs, but they are available upon request. The low reliability on these constructs likely stems from the limited number of items we were able to use due to survey space constraints, as well as the difficulty younger students had in grasping the meanings of the items, especially those which were reverse-coded.

We present results in our main findings from one item from the emotional empathy construct where students answered the extent to which they agreed or disagreed with the statement "I want to help people who are treated badly." This question has high readability and is part of a larger empathy scale adapted for use with younger children from the Interpersonal Reactivity Index (Davis, 1980), a well-validated empathy measure for adults (Garton & Gringart, 2005). However, results for this item should be interpreted cautiously given that it is only a single item. Finally, college aspirations were captured by a single item ("I plan to go to college") and are indicated by a binary measure of whether students strongly agreed or not. We also provide results for the school engagement and cognitive empathy individual survey items, as well as item sources, in Appendix Table A2.

Sample

There were 15,886 students in 3rd through 8th grades assigned to treatment (intent to treat) and control. We provide student demographics, as well as baseline measures of achievement, attendance, and discipline in Table 1. The vast majority of the students in our sample (86 percent) qualified for FRL. HISD records also contain an indicator for whether students were from households designated as living in poverty based on Texas Education Agency's Public Education Information Management System. This variable indicates whether a student's household has an income at or below the federal poverty line, is eligible for Temporary Assistance to Needy Families or

Table 1. Descriptive statistics by treatment status.

Variable	Treatment	Control	Difference
Grade level	4.980	5.175	-0.196
	(0.462)	(0.473)	(0.658)
Female	0.482	0.490	-0.008
	(0.006)	(0.008)	(0.009)
FRL Status:			
Not FRL	0.142	0.141	0.001
	(0.039)	(0.030)	(0.048)
Reduced lunch eligible	0.069	0.071	-0.002
	(0.009)	(0.009)	(0.013)
Free lunch eligible (non-poverty)	0.311	0.300	0.011
	(0.023)	(0.028)	(0.036)
Free lunch eligible (poverty)	0.478	0.488	-0.010
	(0.040)	(0.031)	(0.050)
Race/Ethnicity:			
Black	0.246	0.255	-0.010
	(0.063)	(0.067)	(0.092)
Hispanic	0.686	0.679	0.007
	(0.069)	(0.066)	(0.095)
White	0.028	0.035	-0.007
	(0.016)	(0.017)	(0.023)
Special education	0.085	0.089	-0.004
	(0.007)	(0.010)	(0.012)
English language learner (ELL)	0.340	0.323	0.017
	(0.052)	(0.045)	(0.069)
Prior year disciplinary infraction	0.064	0.096	-0.033
	(0.021)	(0.027)	(0.035)
Baseline math	0.005	-0.089	0.094
	(0.090)	(0.077)	(0.118)
Baseline reading	0.029	-0.162	0.191
	(0.110)	(0.072)	(0.130)
Prior year attendance rate	0.969	0.965	0.004
	(0.003)	(0.004)	(0.004)
Sample Size	8,034	7,852	

Notes: No differences found to be statistically significant at p-value < 0.05; standard errors in parentheses adjusted for school-level clustering. State standardized math and reading scores are standardized with a mean of zero and a standard deviation of one. Baseline test scores provided for students enrolled in grades 4 through 8; differences in demographic variables, by treatment status, are also not statistically different from zero for our grade 4 through 8 preferred analytic sample. A joint *F*-test from a model regressing the treatment indicator on the full list of covariates failed to reject the null hypothesis that the covariates are jointly equal to zero (p-value = 0.25).

other public assistance, received a Pell Grant or comparable state program of need-based financial assistance, is eligible for programs assisted under Title II of the Job Training Partnership Act, or is eligible for food stamps. In terms of race/ethnicity, 68 percent of students identified as Hispanic, 25 percent as Black, and 3 percent as White. Thirty-three percent of students in our sample were receiving ELL services.

We do not find any statistically significant differences across baseline observables for treatment and control students, though there are some visible differences in baseline math and reading achievement. Defining our preferred analytic sample involves a tradeoff between maximizing our sample size versus controlling for prior achievement, which limits us to students enrolled in 4th through 8th grades because standardized tests are first administered in 3rd grade. Because this is an RCT, estimates using the full sample have strong internal validity, but including baseline covariates for student achievement can increase precision and correct for treatment/control imbalances due to chance (Gerber & Green, 2012). Thus, we provide results for the unrestricted sample that includes 3rd graders, as well as our preferred analytic sample of the 4th through 8th graders that adds controls for prior test score achievement. All of our models also include a complete set of demographic and control variables, which we describe in the next section. Finally, of the 12,330 students in 4th through 8th grades enrolled in AAI treatment and control schools at the time of randomization, 1,426 (12 percent) and 1,613 (13 percent) were missing baseline reading and math scores, respectively. We generated baseline scores for these students using 50 imputed datasets via multiple imputation by chained equations.

Excluding students who ordinarily receive testing accommodations left us with 14,265 potential students for our survey sample. An additional 1,137 students could not be included because they graduated out of AAI treatment or control group schools (scheduled school transfers for the first cohort), and 373 students left their schools prior to survey data collection (unscheduled transfers). These exclusions left us with a potential survey sample of 12,755 students in 3rd through 8th grades and 9,749 students in 4th through 8th grades with baseline test scores, our preferred analytical sample. Survey exclusion and school transfer rates were not statistically different for treatment versus control schools.

We successfully collected and linked outcome survey data to the district's administrative data for 10,066 eligible 3rd through 8th grade students (79 percent), and 7,640 eligible 4th through 8th grade students (78 percent of the sample with baseline test scores). We do not find any statistically significant differences across baseline observables for survey-eligible treatment and control students or for those who participated in the survey (Appendix Table A3). These are high response rates for original data collection efforts from a historically underserved population. However, a greater percentage of control group students completed the survey (7.3 percentage points for the 3rd through 8th grade sample; 7.7 percentage points for the 4th through 8th grade sample), though this difference in survey participation was not statistically significant. We address this concern through multiple robustness checks in our results section.

ANALYSIS

The experimental design of this study provides a straightforward analytical strategy. We estimate the AAI intent-to-treat effects with the following model:

$$Y_{ism} = \alpha + \beta AAI_s + Match_m\delta + X_i\gamma + \varepsilon_{ism}. \quad (1)$$

Where Y signifies an outcome of interest for student i, enrolled in school s, and matched pair m. *AAI* is a dichotomous variable that indicates whether the student was enrolled in a school that was randomly assigned to participate in the AAI; *Match* is a vector of dummy variables for the pre-randomization matched pairs; X is a vector of baseline student demographics, specifically student grade level, gender, race/ethnicity, FRL eligibility and poverty status (we include this indicator to improve the precision of our socioeconomic status measure by distinguishing the 44 percent of FRL-eligible students living above the federal poverty line to the

56 percent living at or below it), an indicator for receipt of special education services, and ELL status. In our preferred models restricted to 4th through 8th grade students, we include prior year absences, an indicator for having received a prior year disciplinary infraction, and prior year math and reading scores; the error term ε is clustered at the school level.

We also conduct exploratory analyses to examine whether the AAI treatment produces heterogeneous effects across subgroups of policy interest. Previous studies have found that effects of arts education interventions tend to be more pronounced with students from historically-marginalized communities, who tend to have less access to arts learning opportunities (Catterall, et al., 2012; Greene et al., 2014; Podlozny, 2000). We run separate analyses based on elementary or middle school status, gender, race/ethnicity, socioeconomic status, and ELL status. We also examine elementary ELL students as a distinct subgroup, as we have a large sample of elementary-aged ELL students, the intervention's primary focus was on elementary schools, and prior research suggests they may respond more to arts-based learning (Peppler et al., 2014). These analyses provide preliminary information on policy-relevant heterogeneous effects that can guide future research. Moreover, because we examine 11 subgroups, we conduct adjustments for multiple comparisons within each outcome domain and identify any cases where the false discovery rate suggests the finding may be due to chance in table notes (Benjamini & Hochberg, 1995). We subjected our main findings to multiple comparison testing; these results were unaffected and are provided in Appendix Table A4 along with a detailed description of the procedure.

In addition to our standard sampling-based inference approach, we also test the likelihood of obtaining estimated treatment effects using randomization-based statistical inference because we have a relatively small number of clusters. This procedure allows for inference conditional on the randomization stratification. We reassign treatment status within matched pairs, and estimate "treatment effects" for each outcome with 1,000 simulated random draws, generating p-values that are the proportion of instances where simulated treatment effects are greater in absolute value than our treatment effect estimate. The full and subgroup results are presented in Appendix Tables A4, A5, and A6.

RESULTS

Results for the full sample are provided in Table 2. We find three statistically significant results for the full sample using our preferred model (column 4). Increasing students' arts educational experiences reduces the proportion of students receiving a disciplinary infraction by 3.6 percentage points; increases writing achievement by 0.13 of a standard deviation; and increases students' emotional empathy for others by 0.07 of a standard deviation. The results are generally consistent across different specifications for all outcomes, with the exception of effects on reading achievement, which are not significant in our preferred models that control for prior achievement. The reading result exception is consistent with the inclusion of baseline measures improving our model due to some small differences in our treatment and control groups due to chance. When we generate p-values using randomization-based statistical inference, our interpretation of the writing achievement effect and emotional empathy remains qualitatively the same (Appendix Table A4). However, the effect on discipline ($p = 0.14$) falls shy of statistical significance. As such, we have somewhat less confidence in the statistical significance for this outcome.

Using our preferred model, exploratory analysis of subgroup effects on non-survey outcomes finds the overall reduction in students receiving disciplinary infractions is present for all groups except elementary and ELL subgroups (Table 3). Effects are

Table 2. Arts Access Initiative treatment effects.

Outcome	(1)	(2)	(3)	(4)
	-0.042***	-0.042***	-0.051***	-0.036***
Discipline	(0.012)	(0.012)	(0.013)	(0.008)
	15,886	15,886	12,330	12,330
	0.003	0.002	0.002	-0.001
Attendance rate	(0.002)	(0.002)	(0.002)	(0.001)
	15,886	15,886	12,330	12,330
	0.098*	0.099*	0.078*	-0.019
Reading	(0.046)	(0.041)	(0.036)	(0.020)
	14,628	14,628	11,343	11,343
	0.064	0.058	0.042	-0.001
Math	(0.058)	(0.051)	(0.051)	(0.035)
	14,576	14,576	11,300	11,300
	0.217**	0.203**	0.203**	0.128**
Writing	(0.066)	(0.059)	(0.059)	(0.043)
	4,842	4,842	4,842	4,842
	0.074	0.081	0.081	-0.023
Science	(0.070)	(0.065)	(0.065)	(0.057)
	4,597	4,597	4,597	4,597
	0.063	0.072	0.083	0.080
School engagement	(0.061)	(0.053)	(0.056)	(0.056)
	10,064	10,064	7,640	7,640
	0.030	0.032	0.020	0.012
College aspirations	(0.023)	(0.023)	(0.019)	(0.018)
	10,028	10,028	7,620	7,620
	0.068*	0.075**	0.090**	0.072**
Emotional empathy	(0.025)	(0.025)	(0.028)	(0.025)
	9,987	9.987	7,590	7,590
	0.035	0.042	0.052	0.039
Cognitive empathy	(0.044)	(0.040)	(0.041)	(0.039)
	9,933	9,933	7,552	7,552
Matched pair FEs	X	X	X	X
Demographics		X	X	X
Restricted to grades 4–8			X	X
Prior performance				X

Notes: *** statistically significant (two-tailed) at $p < 0.001$; ** at $p < 0.01$; * at $p < 0.05$; standard errors in parentheses clustered at the school-level. Test scores and survey outcomes are standardized with a mean of zero and a standard deviation of one. Demographic control variables include grade-level indicators, gender, an indicator for receiving English language learning services, special education status, free- and reduced-price lunch and poverty statuses, and race/ethnicity. Prior performance control variables are previous year's math and reading scores, attendance rate, and an indicator for whether a student had received a disciplinary infraction. There were no missing demographic data; missing prior achievement data imputed with multiple imputation with chained equations using 50 imputations. No significant outcomes fail adjustments for multiple comparisons.

notably larger for middle school, male, and Black students. Writing impacts are consistently positive across subgroups, with the exception of middle schoolers, Black students, and FRL-non-poverty students; FRL non-poverty students fall just shy of significance (p-value = 0.051). Writing effects are notably larger for ELL students, especially in elementary grades.

When exploring survey-based outcomes across subgroups, the positive effect on emotional empathy is largely driven by elementary, female, non-FRL, and ELL students (Table 4). We find additional positive impacts on survey measures for which

Table 3. Subgroup treatment effects on behavioral and test score outcomes.

Population	Discipline	Attendance	Reading	Math	Writing	Science
	-0.036***	-0.001	-0.019	-0.001	0.128**	-0.023
Overall	(0.008)	(0.001)	(0.020)	(0.035)	(0.043)	(0.057)
	12,330	12,330	11,343	11,300	4,842	4,597
	-0.006	-0.001	-0.005	0.027	0.197**	0.034
Elementary	(0.006)	(0.001)	(0.035)	(0.038)	(0.063)	(0.051)
	6,428	6,428	5,934	5,924	3,107	2,822
	-0.068***	-0.001	-0.023	-0.050	0.050	-0.103
Middle	(0.003)	(0.002)	(0.025)	(0.061)	(0.034)	(0.115)
	5,902	5,902	5,409	5,376	1,735	1,775
	-0.026***	-0.001	-0.008	0.005	0.149**	0.037
Female	(0.007)	(0.001)	(0.022)	(0.049)	(0.045)	(0.053)
	5,989	5,989	5,559	5,539	2,345	2,278
	-0.046***	-0.001	-0.020	-0.007	0.130*	-0.022
Male	(0.012)	(0.001)	(0.024)	(0.030)	(0.052)	(0.053)
	6,341	6,341	5,784	5,761	2,497	2,319
	-0.050***	-0.002	-0.061	0.054	0.050	0.063
Black	(0.012)	(0.003)	(0.040)	(0.053)	(0.058)	(0.065)
	3,013	3,013	2,673	2,672	1,191	1,090
	-0.032**	-0.001	-0.004	-0.032	0.153*	-0.018
Hispanic	(0.010)	(0.001)	(0.022)	(0.038)	(0.057)	(0.074)
	8,564	8,564	7,978	7,940	3,334	3,195
	-0.044*	-0.003*	0.042	0.040	0.240*	-0.081
Not FRL	(0.017)	(0.001)	(0.048)	(0.081)	(0.094)	(0.101)
	1,731	1,731	1,604	1,594	688	696
	-0.029***	-0.001	-0.022	-0.058	0.118	0.021
FRL-non-poverty	(0.008)	(0.001)	(0.022)	(0.042)	(0.058)	(0.060)
	4,703	4,703	4,375	4,347	1,815	1,759
	-0.038***	-0.001	-0.027	0.019	0.119**	0.015
FRL-poverty	(0.009)	(0.002)	(0.025)	(0.035)	(0.044)	(0.046)
	5,896	5,896	5,364	5,359	2,339	2,142
	-0.020	0.002	0.005	-0.001	0.271**	0.052
ELL	(0.009)	(0.001)	(0.027)	(0.046)	(0.093)	(0.056)
	3,726	3,726	3,463	3,423	1,621	1,281
	-0.001	0.002	0.014	0.038	0.348**	0.028
Elementary ELL	(0.005)	(0.001)	(0.037)	(0.035)	(0.112)	(0.069)
	2,440	2,440	2,288	2,280	1,271	1,012

Notes: *** statistically significant (two-tailed) at $p < 0.001$; ** at $p < 0.01$; and * at $p < 0.05$; standard errors in parentheses clustered at the school-level. Number of student observations provided below standard errors. All models include controls for matched pair, demographics, and prior performance. Test scores are standardized with a mean of zero and a standard deviation of one. No significant outcomes fail adjustments for multiple comparisons.

we did not detect overall effects, particularly for elementary and ELL student subgroups. In addition to these subgroups exhibiting more-pronounced effects on emotional empathy, elementary-aged students demonstrate positive treatment effects on school engagement, college aspirations, and cognitive empathy; ELL students similarly have positive effects on college aspirations and cognitive empathy. Results are particularly pronounced for ELL students in elementary grades. It is also worth noting that there is evidence of negative impacts with middle schoolers on school engagement and college aspirations. These negative findings may be related to the less robust implementation of the program in middle schools, which we consider further in the discussion section.

Table 4. Subgroup treatment effects on survey outcomes.

Population	School engagement	College aspirations	Emotional empathy	Cognitive empathy
	0.080	0.012	0.072**	0.039
Overall	(0.056)	(0.018)	(0.025)	(0.039)
	7,640	7,620	7,587	7,552
	0.211***	0.052*	0.129***	0.095*
Elementary	(0.058)	(0.024)	(0.032)	(0.046)
	4,622	4,609	4,591	4,568
	-0.125*	-0.046**	-0.000	-0.031
Middle	(0.048)	(0.011)	(0.016)	(0.052)
	3,018	3,011	2,999	2,984
	0.105	0.017	0.082**	0.060
Female	(0.065)	(0.019)	(0.030)	(0.041)
	3,834	3,826	3,816	3,798
	0.056	0.003	0.054	0.010
Male	(0.054)	(0.023)	(0.040)	(0.042)
	3,806	3,794	3,774	3,754
	0.119	-0.039	0.067	-0.089
Black	(0.071)	(0.027)	(0.045)	(0.054)
	1,676	1,673	1,662	1,650
	0.050	0.036	0.067*	0.065
Hispanic	(0.063)	(0.022)	(0.032)	(0.043)
	5,320	5,307	5,289	5,262
	0.095	-0.039	0.114*	0.094
Not FRL	(0.072)	(0.033)	(0.050)	(0.098)
	1,054	1,050	1,048	1,042
	0.096	0.030	0.071**	0.091*
FRL-non-poverty	(0.063)	(0.023)	(0.023)	(0.044)
	2,953	2,947	2,934	2,920
	0.059	0.011	0.063	-0.024
FRL-poverty	(0.060)	(0.020)	(0.035)	(0.035)
	3,633	3,623	3,608	3,590
	0.143	0.065*	0.157***	0.113*
ELL	(0.081)	(0.027)	(0.042)	(0.052)
	2,618	2,612	2,603	2,592
	0.241**	0.096***	0.180***	0.134*
Elementary ELL	(0.081)	(0.026)	(0.049)	(0.055)
	1,958	1,952	1,946	1,937

Notes: *** statistically significant (two-tailed) at $p < 0.001$; ** at $p < 0.01$; and * at $p < 0.05$; standard errors in parentheses adjusted for school-level clustering. Number of student observations provided below standard errors. Adjustments for multiple comparisons suggest that the significant effect for the Hispanic subgroup on emotional empathy and all subgroup effects on cognitive empathy may be false discoveries and should be interpreted with caution.

Additional Robustness Checks with Survey Data

Although we have little missing data from administrative sources, our survey response rates, while high, still leave room for nonresponse bias. Though the respondent sample is balanced across observable characteristics, students in the control group had a higher survey completion rate (by 7 percentage points), though this difference was not statistically significant. As a result, we subject our survey measures to multiple additional robustness checks (Table 5). For our first robustness check, we fully impute missing outcome data for the survey-eligible grade 4 through 8 sample. This full sample imputation analysis has all the advantages of our

Table 5. Robustness checks for survey outcomes.

Outcome	Preferred analysis	Full Sample imputation	IPWRA	Pre-Post
	0.080	0.083	0.061**	0.115**
School engagement	(0.056)	(0.046)	(0.023)	(0.038)
	7,640	9,749	7,640	5,763
	0.012	0.009	0.010	0.075
College aspirations	(0.018)	(0.015)	(0.011)	(0.045)
	7,620	9,749	7,620	5,704
	0.072**	0.078**	0.066**	0.095**
Emotional empathy	(0.025)	(0.023)	(0.024)	(0.028)
	7,590	9,749	7,590	5,611
	0.039	0.037	0.031	0.059
Cognitive empathy	(0.039)	(0.032)	(0.025)	(0.036)
	7,552	9,749	7,552	5,584

Notes: *** statistically significant (two-tailed) at $p < 0.001$; ** at $p < 0.01$; and * at $p < 0.05$; standard errors in parentheses clustered at the school-level. Number of student observations provided below standard errors. Pre-post analysis controls for baseline measures of the survey outcome for cohort 2; cohort 1 students had already received the treatment prior to the initial survey implementation and were, therefore, excluded from this analysis. For full-sample imputation, missing outcome survey responses, in addition to missing baseline achievement were imputed for survey-eligible students.

preferred model with the full set of baseline controls and has the effect of increasing the sample size by roughly 2,100 students (column 2). The results of this approach are qualitatively similar to our main models, though are slightly more precisely estimated. Next, we address potential baseline imbalance across the treatment and control groups by conducting an inverse-probability-weighted regression adjustment (IPWRA) for the survey-respondent sample (column 3). The IPWRA estimator uses the parameters of the model to compute the inverse probability of being in the treatment group, then applies these weights to compute weighted regression coefficients to predict average treatment effects. This method yields results similar to our main models, with a positive significant effect on emotional empathy, and also generates a significant positive effect on school engagement that is more precisely estimated.

Finally, it would be preferable to control for outcome measures collected at baseline to adjust for any pre-treatment differences and improve the precision of our estimates, as we do in our analysis of test score outcomes, discipline, and attendance; these baseline items are also used as controls in all our survey outcome models. Fortunately, the availability of baseline survey measures for students in the second cohort allow us to do this, albeit with a smaller sample (column 4). Similar to the IPWRA estimates, controlling for survey outcomes collected at baseline generates positive significant effects on school engagement and emotional empathy. The significant estimate of 0.12 for school engagement increases in magnitude by 44 percent compared to our main model, whereas the significant effect of 0.10 on emotional empathy increases by roughly 32 percent. Though we cannot fully rule out nonresponse bias on unobservable characteristics, these approaches serve as strong robustness checks of our estimates. Taken together, the preponderance of evidence from our sensitivity tests suggests our preferred survey results are valid causal estimates.

Moreover, because the focus of the intervention, our pattern of results, and prior research suggest that elementary and ELL students in particular benefited from the AAI, the robustness of these findings has particular policy relevance. Therefore, we

subject these subgroup findings to the same survey robustness checks as our full sample (Appendix Tables A7 through A9). The results of these sensitivity tests largely confirm the integrity of the positive findings for elementary and ELL students on school engagement, college aspirations, emotional empathy, and cognitive empathy.

DISCUSSION & CONCLUSION

Our investigation is the first large-scale randomized control trial of an arts education program implemented in an authentic school setting. We find that increases in students' arts learning experiences significantly improved policy-relevant educational outcomes in a diverse array of elementary and middle schools in the nation's 7th largest school district. Students in the treatment group received fewer disciplinary infractions, demonstrated increases in writing achievement, and expressed greater emotional empathy. Moreover, when we restrict our analysis to elementary students, who were the main focus of the intervention, we find additional positive effects on school engagement, college aspirations, and cognitive empathy. These results demonstrate that the arts positively affect meaningful educational outcomes and provide important policy considerations for strategies to retain arts education in under-resourced schools (Bowen & Kisida, 2017; Catterall et al., 2012; Deasy, 2002; DiMaggio, 1982; Fiske, 1999; Ruppert, 2006).

Though we find no effects on math, reading, or science achievement, the positive effects on writing achievement are particularly noteworthy. Despite many dubious claims that arts education has transfer effects on tested subjects, there is little theoretical basis or rigorous evidence supporting such an expectation (Winner et al., 2013). The most likely pathway for arts education to improve test scores in other subjects would be spillover effects from improvements in school engagement. Writing achievement, however, is more closely aligned with aspects of arts learning than other tested subjects. Many of the programs students participated in included opportunities for self-expression and reflection, and a number of programs included student writing exercises either through a specific focus on literary arts or arts integration programs that included writing. The STAAR test assesses students' writing proficiency through two components—multiple choice items on mechanical skills and open-response expository essays to assess composition skills. When we disaggregate student scores on this assessment, we find significant increases for treatment group students on both sections, but effects are greater in magnitude and significance for the written composition portion of the exam (0.18 SD; $p < 0.001$) than the writing mechanics section (0.09 SD; $p = 0.02$). This finding aligns with the theory that students became better at expressing themselves and articulating their own ideas through participation in arts experiences.

Positive effects on social-emotional learning outcomes are also especially noteworthy, as a growing body of evidence demonstrates their importance for later-life outcomes (Jackson, 2018, 2020). Though there is a strong theoretical basis that the nature of arts education could affect social-emotional outcomes, thus far there has been little rigorous evidence (American Academy of Arts and Sciences, 2021). Analysis on our full sample finds positive effects on emotional empathy, while students in elementary schools also demonstrated large positive effects on school engagement and cognitive empathy. The positive effects we observe on social-emotional outcomes are also well-aligned with our findings of reduced disciplinary infractions, as well as higher college aspirations for elementary school students. In related work, research finds that schools' contributions to social-emotional development leads to fewer disciplinary infractions, fewer school-based arrests, and higher rates of college attendance and persistence (Jackson, 2020). Arts education presents a promising

option for policymakers interested in improving social-emotional learning outcomes and student behavior.

Many of our main findings are consistent across student subgroups; however, there were notable variations that could inform policy decisions and future research efforts. First, results were generally more pronounced for elementary-level students. In addition to demonstrating more-pronounced effects on writing achievement and emotional empathy, these students experienced positive treatment effects on school engagement, college aspirations, and cognitive empathy. Not only did results for elementary students tend to be more pronounced, but they were more robust to alternative specifications, including randomization-based inference tests and the multiple robustness tests we applied to our survey outcomes (Appendix Tables A6 and A7). We expect this is largely attributable to elementary schools being the primary focus during the first two years of the AAI, comprising 86 percent of our sample. Elementary schools were also less likely to have full-time arts specialists prior to the intervention, and the disproportionate attention given to these schools likely improved fidelity of implementation. Based on AAI administrative records, elementary schools were more likely to receive the deeper, sustained in-school artist residencies that lasted entire semesters and covered entire grade levels. It could also be the case that younger students are more receptive to arts education experiences, as educational interventions tend to have greater effects in early years (Heckman, 2006).

Similar to elementary-level students, ELL and elementary-ELL students demonstrated more pronounced effects on school engagement, college aspirations, and empathy, and their results were similarly robust to multiple alternative specifications. Research and theory suggests ELL students may particularly benefit when arts-learning techniques are integrated to deliver core content. Related studies have shown a consistent relationship between arts integration and increases in academic outcomes for ELL students, including increases in written and oral language skills, increased student engagement, and reductions in absences for ELL students (Brouillette et al., 2014; Greenfader et al., 2015; Peppler et al., 2014). Authors of these studies suggest that arts learning increases verbal interactions between students and teachers and offers multiple pathways to connect with educational content, both areas where ELL students may be at a disadvantage in traditional educational settings. Moreover, the AAI programming tended to have a strong emphasis on art from a diverse array of cultures, which may have been especially engaging for students who do not identify with the dominant culture.

We also find some puzzling results that do not align with our initial expectations. For example, we find no effects on student attendance across the main sample or any subgroup, even for subgroups who report increased school engagement. We suspect that this is related to the fact that the bulk of our students are in their earlier years of schooling, and are still too young to have the agency to make their own decisions about whether or not to attend school. We also found somewhat inconsistent patterns suggesting that historically-underserved students responded more to the intervention. Though the effects for ELL students are consistent with this hypothesis, we generally observe similar effects across different indicators of socioeconomic status. It is worth noting, however, that nearly 90 percent of our sample is FRL-eligible, so our sample may lack the socioeconomic diversity to investigate this definitively. Additionally, we find negative effects for middle-schoolers on school engagement and college aspirations. We expect this is due to the lesser focus on middle schools and the generally lower treatment dosage. Elementary schools had lower levels of arts indicators at baseline, including some who lacked any full-time arts specialists, and so gravitated toward programs that provided more broad coverage. Programming in middle schools tended to be more piecemeal one-off experiences. As a result, smaller proportions of middle school students were exposed to the programming. This programming not only diluted the dosage, but may have generated a negative

expectations bias or resentful demoralization, two construct validity threats that would be more likely to show up in survey outcomes (Shadish et al., 2001). Finally, it is possible that the results are merely spurious, as neither comes close to significance when using randomization-based inference. These complexities regarding who most benefits and under what conditions remain ripe areas for future research.

Our study is not without limitations. While random assignment to the AAI treatment allows us to confidently infer a causal relationship between the treatment and assessed outcomes, we remain less certain about the extent to which these results would be achieved in other contexts. In order to be considered for participation in the AAI, principals had to willingly commit to improve their schools' arts educational offerings. While our research design ensures their motivation is balanced across the treatment and control groups, our results may not be generalizable to schools lacking leaders dedicated to supporting the arts. Moreover, Houston is an arts-rich metropolis with a wealth of cultural institutions and organizations committed to school partnerships. Smaller or less-arts-rich cities, as well as suburban and rural communities, may lack such resources. However, it is worth noting that there are widespread instances of school-community partnerships with arts organizations across communities of various sizes (Bowen & Kisida, 2017). It is also the case that the majority of schools participating in the AAI had experienced pressure from test-based accountability sanctions, which tend to narrow educational offerings while intensifying the emphasis on student standardized test achievement (Booher-Jennings, 2005; Deming et al., 2016; Neal, 2010; Neal & Schanzenbach, 2010; West, 2007). As mentioned previously, the AAI initially recruited schools with the lowest levels of arts resources. As a result, the counterfactual condition for treatment group schools were control group schools with severe deficiencies in arts educational offerings. This context may have led to a situation ripe for treatment group schools to outperform the control group as a result of the intervention, but a similar program in schools with higher initial levels of arts resources may not produce the same effects.

Though we can state with confidence that our findings are a result of increasing arts educational experiences, we do not know which aspects of the treatment were most responsible for influencing outcomes. The arts experiences that students received were incredibly diverse in terms of program format and arts disciplines. The AAI director and staff worked with principals to formulate plans and program selections that would align with the goals they had for their schools, which typically resulted in students receiving dance, music, theatre, and visual arts educational experiences through teaching-artist residencies, professional artist workshops and performances on- and off-campus, before- or after- school programs, and field trips. Due to the variation and endogeneity in program selection across schools, we cannot discern which particular types of offerings mattered most. Finally, our analysis is restricted to shorter-term outcomes, and it remains to be seen if these effects will ultimately compound, plateau, or diminish over time. Critical next steps in this field of study will be to examine whether particular formats and varieties of arts educational programs are successful in sustaining effects over longer periods of time.

Accountability systems and resource constraints have led to a narrowing of educational offerings and a decline in arts education. Despite the logic behind narrowing offerings, we find no evidence that increasing arts educational experiences is detrimental to student growth in tested subjects. At the same time, we find evidence that restoring the role of arts in education generates significant gains in other important domains. This study provides critical evidence that students' exposure to arts educational opportunities positively impacts meaningful, policy-relevant outcomes. Our findings are particularly timely, as policymakers and the public are increasingly turning their attention to education reforms that improve school engagement, school climate, and other social-emotional and behavioral outcomes. Education

policymakers should consider these benefits when assessing the opportunity costs that occur when arts education is decreased or eliminated.

DANIEL H. BOWEN is Associate Professor in the School of Education & Human Development at Texas A&M University and Co-Director of the Arts, Humanities, & Civic Engagement Lab, 532 Harrington Tower, 4222 TAMU, Texas A&M University, College Station, TX 77843 (email: danielhbowen@gmail.com).

BRIAN KISIDA is Assistant Professor at the Truman School of Government & Public Affairs at the University of Missouri and Co-Director of the Arts, Humanities, & Civic Engagement Lab, 615 Locust Street, Columbia, MO, 65211 (email: briankisida@gmail.com).

ACKNOWLEDGMENTS

We thank the Houston Endowment, National Endowment for the Arts (#17-3800-7011), and the Spencer Foundation (#201600161) for supporting this research, as well as the Houston Independent School District, participating schools, the Houston Education Research Consortium, Young Audiences Inc. of Houston, and the participating arts partners. We thank the following individuals who were instrumental to this study: Richard A. Carranza, Daniel Gohl, Andrew Houlihan, Grenita Lathan, Deborah Lugo, Wenden Sanders, Adam Stephens, Carla Stevens, and Douglas Torres-Edwards. Finally, we acknowledge the valuable research assistance of Holly Heard, Brenna Lin, Jean Madsen, and Kelsey Ragan, as well as the helpful feedback we received from Shaun M. Dougherty, Ruth N. López Turley, Marcus A. Winters, and our editor and anonymous reviewers.

REFERENCES

American Academy of Arts and Sciences. (2021). Art for life's sake: The case for arts education. https://www.amacad.org/project/commission-arts

Austin Independent School District. (2016). *2015–16 student climate survey*. https://www.austinisd.org/sites/default/files/dre-surveys/15.53RBa_Student_Climate_Survey_Results_Summary_for_Spring_2014_Through_Spring_2016.pdf

Balfanz, R., Legters, N., West, T. C., & Weber, L. M. (2007). Are NCLB's measures, incentives, and improvement strategies the right ones for the nation's low-performing high schools? American Educational Research Journal, 44(3), 559–593. https://doi.org/10.3102/0002831207306768

Bassok, D., Latham, S., & Rorem, A. (2016). Is kindergarten the new first grade? AERA Open, 2(1), 1–31. https://doi.org/10.1177/2332858415616358

Batel, S. (2017). Measuring success: An overview of new school classification indicator under ESSA. Center for American Progress. https://cdn.americanprogress.org/content/uploads/2017/08/08080020/FifthIndicator-MeasuringSuccess-brief2.pdf

Benjamini, Y., & Hochberg, Y. (1995). Controlling for the false discovery rate: A practical and powerful approach to multiple testing. Journal of the Royal Statistical Society, Series B, 57(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Booher-Jennings, J. (2005). Below the bubble: "Educational triage" and the Texas accountability system. American Educational Research Journal, 42(2), 231–268. https://doi.org/10.3102/00028312042002231

Bourdieu, P. (1977). Cultural reproduction and social reproduction. In J. Karabel and A. Halsey (Eds.), Power and ideology in education (pp. 487–511). Oxford University Press.

Bowen, D. H., Greene, J. P., & Kisida, B. (2014). Learning to think critically: A visual art experiment. Educational Researcher, 43(1), 37–44. https://doi.org/10.3102/0013189x13512675

Bowen, D. H., & Kisida, B. (2017). The art of partnerships: Community resources for arts education. Phi Delta Kappan, 98(7), 8–14. https://doi.org/10.1177/0031721717702624

Brouillette, L., Childress-Evans, K., Hinga, B., & Farkas, G. (2014). Increasing engagement and oral language skills of ELLs through the arts in the primary grades. Journal for Learning through the Arts, 10(1), 1–25. http://doi.org/10.21977/D910119300

Catterall, J. S., Dumais, S. A., & Hampden-Thompson, G. (2012). *The arts and achievement in at-risk youth: Findings from four longitudinal studies*. National Endowment for the Arts.

Center on Education Policy. (2008). Instructional time in elementary schools: A closer look at changes for specific subjects. Arts Education Policy Review, 109(6), 23–28. https://doi.org/10.3200/aepr.109.6.23-28

Cotton, K. (2001). *Developing empathy in children and youth*. School Improvement Research Series Close Up # 13. https://educationnorthwest.org/sites/default/files/developing-empathy-in-children-and-youth.pdf

Da Camera (2022). *Educational Programs*. https://www.dacamera.com/learn/educational-programs/

Davis, M. H. (1980). A multidimensional approach to individual differences in empathy. JSAS Catalog of Selected Documents in Psychology, 10, 85.

Deasy, R. (2002). *Critical links: Learning in the arts and student academic and social development*. Arts Education Partnership.

Dee, T. S., Jacob, B., & Schwartz, N. L. (2013). The effects of NCLB on school resources and practices. Educational Evaluation and Policy Analysis, 35(2), 252–279.

Deming, D. J., Cohodes, S., Jennings, J., & Jencks, C. (2016). School accountability, post-secondary attainment, and earnings. Review of Economics and Statistics, 98(5), 848–862. https://doi.org/10.1162/REST_a_00598

Dewey, J. (1919). Imagination and expression. Teachers College Bulletin, 10(10), 7–15.

DiMaggio, P. (1982). Cultural capital and school success: The impact of status culture participation on the grades of US high school students. American Sociological Review, 47(2), 189–201. https://www.jstor.org/stable/2094962

Duckworth, A. L., Peterson, C., Matthews, M. D., & Kelly, D. R. (2007). Grit: Perseverance and passion for long-term goals. Journal of Personality and Social Psychology, 92(6), 1087–1101. https://doi.org/10.1037/0022-3514.92.6.1087

Education Commission of the States. (2019). *Arts in core academic subjects, assessment and accreditation*. https://c0arw235.caspio.com/dp/b7f930002494fddfe8bf43e1a269

Elpus, K. (2013). Is it the music or is it selection bias? A nationwide analysis of music and non-music students' SAT scores. Journal of Research in Music Education, 61(2), 175–194. https://doi.org/10.1177/0022429413485601

Erickson, H. H., Watson, A. R., & Greene, J. P. (2022). An experimental evaluation of culturally enriching field trips. *The Journal of Human Resources*. Advance online publication. https://doi.org/10.3368/jhr.1020-11242R1

Farkas Duffett Research Group. (2012). *Learning less: Public school teachers describe a narrowing curriculum*. https://www.americansforthearts.org/sites/default/files/cc-learning-less-mar12.pdf

Fiske, E. (Ed.) (1999). *Champions of change: The impact of the arts on learning*. The Arts Education Partnership and the President's Committee on the Arts and Humanities.

Fotofest. (2022). *Literacy through photography*. https://fotofest.org/learningprogram

Gadsden, V. L. (2008). The arts and education: Knowledge generation, pedagogy, and the discourse of learning. Review of Research in Education, 32, 29–61. https://www.jstor.org/stable/20185112

Garton, A. F., & Gringart, E. (2005). The development of a scale to measure empathy in 8- and 9-year old children. Australian Journal of Education and Developmental Psychology, 5, 17–25.

Gerber, A. S., & Green, D. P. (2012). Field experiments: Design, analysis, and interpretation. W.W. Norton & Company.

Gibson, C. (2016). *Dancing to the top: How collective action revitalized arts education in Boston*. EdVestors & Boston Public Schools Arts Expansion Partners. www.edvestors.org/wp-content/uploads/2016/05/BPS-ArtsExpansion-Case-Study.pdf

Government Accountability Office. (2009). Access to arts education: Inclusion of additional questions in education's planned research would help explain why instruction time has decreased for some students. GAO-09-286. https://www.gao.gov/products/gao-09-286

Greene, J. P., Erickson, H. H., Watson, A. R., & Beck, M. I. (2018). The play's the thing: Experimentally examining the social and cognitive effects of school field trips to live theater performances. Educational Researcher, 47(4), 246–254. https://doi.org/10.3102/0013189x18761034

Greene, J. P., Kisida, B., & Bowen, D. H. (2014). The educational value of field trips: Taking students to an art museum improves critical thinking skills, and more. Education Next, 14(1), 78–87.

Greenfader, C. M., Brouillette, L., & Farkas, G. (2015). Effect of a performing arts program on the oral language skills of young English learners. Reading Research Quarterly, 50(2), 185–203. https://doi.org/10.1002/rrq.90

Hardiman, M. M., JohnBull, R. M., Carran, D. T., & Shelton, A. (2019). The effects of arts integrated instruction on memory for science content. Trends in Neuroscience and Education, 14, 25–32. https://doi.org/10.1016/j.tine.2019.02.002

Heckman, J. J. (2006). Skill formation and the economics of investing in disadvantaged children. Science, 312(5782), 1900–1902. https://doi.org/10.1126/science.1128898

Jackson, C. K. (2018). What do test scores miss? The importance of teacher effects on non-test score outcomes. Journal of Political Economy, 126, 1–35. https://doi.org/10.3386/w22226

Jackson, C. K., Porter, S. C., Easton, J. Q., Blanchard, A., & Kiguel, S. (2020). School effects on socioemotional development, school-based arrests, and educational attainment. AER: Insights, 2(4), 491–508. https://doi.org/10.1257/aeri.20200029

Kinney, D. W., & Forsythe, J. L. (2005). The effects of the arts IMPACT curriculum upon student performance on the Ohio fourth-grade proficiency test. Bulletin of the Council for Research in Music Education, 164, 35–48.

Kisida, B., Bowen, D. H., & Greene, J. P. (2016). Measuring critical thinking: Results from an art museum field trip experiment. Journal of Research on Educational Effectiveness, 9(sup1), 171–187. https://doi.org/10.1080/19345747.2015.1086915

Kisida, B., Goodwin, L., & Bowen, D. H. (2020). Teaching history through theater: The effects of arts integration on students' knowledge and attitudes. AERA Open, 6, 1–11. https://doi.org/10.1177/2332858420902712

Konrath, S. & Kisida, B. (2021). Does arts engagement increase empathy and prosocial behavior? A review of the literature. In L. N. Hersey and B. Bobick (Eds.) Engagement in the city how arts and culture impact development in urban areas (pp. 7–38). Lexington Books.

Kraft, M. A. (2019). Teacher effects on complex cognitive skills and social-emotional competencies. The Journal of Human Resources, 54(1), 1–36.

Lacoe, J., Painter, G. D., & Williams, D. (2020). Museums as classrooms: The academic and behavioral impacts of "School in the Park." AERA Open, 6, 1–22. https://doi.org/10.1177/2332858420940309

Ludwig, M. J., Boyle, A., & Lindsay, J. (2017). Review of evidence: Arts integration research through the lens of the Every Student Succeeds Act. American Institutes for Research. https://www.wallacefoundation.org/knowledge-center/Documents/Arts-Integration-Research-Every-Student-Succeeds-Act-ESSA.pdf

Martingano, A. J., & Konrath, S. (2022). How cognitive and emotional empathy relate to rational thinking: Empirical evidence and meta-analysis. The Journal of Social Psychology, 162(1), 1–18. https://doi.org/10.1080/00224545.2021.1985415

McCarthy, K. F., Ondaatje, E. H., Zakaras, L., & Brooks, A. (2004). Gifts of the muse: Reframing the debate about the benefits of the arts. RAND Corporation.

Mehta, J. (2017, February 8). *Schools already have good learning, just not where you think*. Education Week. http://blogs.edweek.org/edweek/learning_deeply/2017/02/why_schools_already_have_good_learning_just_not_where_you_think.html?qs=jal+mehta+periphery

Mercury Chamber Orchestra. (2022). *In-school performances*. https://www.mercuryhouston.org/inschool-performances

Meyer, D., Princiotta, D., & Lanahan, L. (2004). *The summer after kindergarten: Children's activities and library use by household socioeconomic status* (NCES 2004–037). U.S. Department of Education, National Center for Education Statistics.

Multicultural Education and Counseling through the Arts. (2022). *In school + outreach program*. https://www.meca-houston.org/in-school.html

Nameless Sound. (2022). *Creative music communities*. https://www.namelesssound.org/programs/workshops

Neal, D. (2010). Aiming for efficiency rather than proficiency. Journal of Economic Perspectives, 24(3), 119–132. https://doi.org/10.1257/jep.24.3.119

Neal, D., & Schanzenbach, D. W. (2010). Left behind by design: Proficiency counts and test-based accountability. The Review of Economics and Statistics, 92(2), 263–283. https://doi.org/10.1162/rest.2010.12318

New York City Department of Education. (2016). *NYC school survey: 2016 student survey*. https://infohub.nyced.org/docs/default-source/default-document-library/2016nycschoolsurveystudentenglish.pdf?sfvrsn=e50ce333_2

Parsad, B., & Spiegelman, M. (2012). Arts education in public elementary and secondary schools: 1999–2000 and 2009–10. National Center for Education Statistics 2012–014.

Peppler, K. A., Powell, C. W., Thompson, N., & Catterall, J. (2014). Positive impact of arts integration on student academic achievement in English language arts. The Educational Forum, 78(4), 364–377. https://doi.org/10.1080/00131725.2014.941124

Perille. (2016, November 29). *A blueprint for successful arts education: Five strategies for investing in students' creative capital*. Education Week. https://www.edweek.org/ew/articles/2016/11/30/a-blueprint-for-successful-arts-education.html

Podlozny, A. (2000). Strengthening verbal skills through the use of classroom drama: A clear link. Journal of Aesthetic Education, 34(3/4), 91–104. https://doi.org/10.2307/3333644

Rabkin, N., & Hedberg, E. C. (2011). *Arts education in America: What the declines mean for arts participation*. *Based on the 2008 Survey of Public Participation in the Arts* [Research Report# 52]. National Endowment for the Arts.

Redford, J., Burns, S., & Hall, L. J. (2018). *The summer after kindergarten: Children's experiences by socioeconomic characteristics* (NCES 2018–160). U.S. Department of Education, National Center for Education Statistics.

Roksa, J., & Potter, D. (2011). Parenting and academic achievement: Intergenerational transmission of educational advantage. Sociology of Education, 84(4), 299–321. https://doi.org/10.1177/0038040711417013

Roscigno, V. J., & Ainsworth-Darnell, J. W. (1999). Race, cultural capital, and educational resources: Persistent inequalities and achievement returns. Sociology of Education, 72(3), 158–78. http://doi.org/10.2307/2673227

Ruppert, S. S. (2006). *Critical evidence: How the arts benefit student achievement*. National Assembly of State Arts Agencies.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2001). Experimental and quasi-experimental designs for generalized causal inference. Houghton-Mifflin Company.

Soland, J., Hamilton, L. S., & Stecher, B. M. (2013). Measuring 21st century competencies: Guidance for educators. RAND Corporation. http://asiasociety.org/files/gcen-measuring21cskills.pdf

Spreng, R. N., McKinnon, M. C., Mar, R. A., & Levine, B. (2009). The Toronto empathy questionnaire: Scale development and initial validation of a factor-analytic solution to multiple empathy measures. Journal of Personality Assessment, 91(1), 62–71. https://doi.org/10.1080/00223890802484381

Thomas, M. K., Singh, P., & Klopfenstein, K. (2015). Arts education and the high school dropout problem. Journal of Cultural Economics, 39(4), 327–339. https://doi.org/10.1007/s10824-014-9238-x

Wan, Y., Ludwig, M. J., & Boyle, A. (2018). *Review of evidence: Arts education through the lens of the Every Student Succeeds Act*. American Institutes for Research. https://files.eric.ed.gov/fulltext/ED591872.pdf

West, M. (2007). Testing, learning, and teaching: The effects of test-based accountability on student achievement and instructional time in core academic subjects. In C. E. Finn Jr. & D. Ravitch (Eds.) Beyond the basics: Achieving a liberal education for all children (pp. 45–61). The Fordham Institute.

Winner, E., & Cooper, M. (2000). Mute those claims: No evidence (yet) for a causal link between arts study and academic achievement. Journal of Aesthetic Education, 34, 11–75.

Winner, E., & Goldstein, T. R. Vincent-Lancrin, S. (2013). Art for art's sake? The impacts of arts education. OECD Publishing.

Writers in the Schools. (2022). *WITS for Schools*. https://witshouston.org/programs/wits-for-schools/

Yee, V. (2014, April 7). Arts education lacking in low-income areas of New York City, report says. *The New York Times*. https://www.nytimes.com/2014/04/07/nyregion/arts-education-lacking-in-low-income-areas-of-new-york-city-report-says.html