THE SOCIOEMOTIONAL BENEFITS OF THE ARTS:

A New Mandate for Arts Education

STEVEN J. HOLOCHWOST
DENNIE PALMER WOLF
WOLFBROWN

KELLY R. FISHER
KERRY O'GRADY
SCIENCE OF LEARNING INSTITUTE
JOHNS HOPKINS UNIVERSITY

THE SOCIOEMOTIONAL BENEFITS OF THE ARTS:

A New Mandate for Arts Education

Steven J. Holochwost
Dennie Palmer Wolf
WolfBrown

Kelly R. Fisher
Kerry O'Grady
Johns Hopkins University

December 15, 2016

TABLE OF CONTENTS

THE SOCIOEMOTIONAL BENEFITS OF THE ARTS:
A New Mandate for Arts Education

EXECUTIVE SUMMARY

I. CONTEXT

What is the mandate for the William Penn Foundation to invest in arts education?

In this report we argue that there is an evolving and compelling mandate for investing in arts education based on the equitable distribution of educational opportunities and how those opportunities can foster not only students' arts learning but also their socioemotional development. Opportunities to engage in socioemotional learning (SEL) result in the acquisition of skills that are critically important for success in school and life, including the ability to manage behavior and make effective decisions, the maintenance of a positive self-concept, and the capacity to interact productively with others. All young people need to develop these skills, but students in Philadelphia placed 'at risk' by poverty have experienced a steady winnowing of experiences and opportunities that might foster socioemotional development. If arts education is to play a role in restoring these opportunities, the Foundation and the providers it funds must develop a deeper understanding of the specific domains of socioemotional development arts programs can nurture, how to design programs that foster these skills in diverse students and contexts, and how to assess whether programs are effectively building students' socioemotional skills.

To begin building this understanding the Foundation engaged WolfBrown, an arts and arts education research firm, to collaborate with Johns Hopkins' Science of Learning Institute in an evaluation of a cohort of its 2015-2016 arts education grantees. After reviewing the literature on the capacity of arts education to foster socioemotional development, we worked with grantees to formulate a set of socioemotional outcome domains that might be unique to the experience of arts education— as opposed to those that could result from other extra-curricular activities, such as athletics or science clubs— but that were *not* likely to be specific to a single arts discipline, such as theater. Based on these domains we formulated the following guiding questions for our evaluation:

1. Can arts education programs foster students' socioemotional development in proximal domains:[1] interest in the arts, tolerance for others' perspectives, or cultural awareness?

2. Can these programs foster development in more distal domains: perseverance, school engagement, growth mindset, or *academic* goal orientation, self-concept, or self-efficacy?

3. Can these programs foster development in domains of self-awareness related to the arts: *artistic* goal orientation, self-concept, or self-efficacy?

---

1   Here and throughout we use the terms "proximal" and "distal" to refer to domains of socioemotional development that are more closely related to the experience of arts education (such as interest in the arts), in the case of proximal domains, or less closely related to those experiences (such as academic self-concept), in the case of distal domains.

II. FINDINGS:
THE IMPACT OF ARTS EDUCATION ON SOCIOEMOTIONAL DEVELOPMENT

These research questions were addressed in two phases. First, we examined whether there were overall or omnibus effects of arts education on students' socioemotional development in proximal and distal domains. However, given the tremendous diversity of programs and the students they served, we did not expect to observe many of these omnibus effects. Rather, we expected that program effects would be contingent upon program factors – such as the intensity of dosage of instruction – and student factors, such as age.

A. Can Programs Foster Socioemotional Development for All Students?

Our results indicated that across a set of programs included in the Foundation's cohort of grantees for the 2015-16 school year, participation led to modest increases in students' interest in the arts. While this finding might be expected, it is notable given the diversity of our sample of programs in terms of artistic discipline (e.g., music), the intensity or dosage of instruction and the diversity of the students those programs served according to factors such as age. It suggests that even relatively-brief, compulsory programs of arts education can kindle students' interest in the arts.

B. Understanding the Diversity of Programs and Students

The next phase of our analyses revealed that all the remaining effects of programs were contingent upon factors related to the diversity of these students, with two factors exerting particularly potent influences:

- Student Age: Arts education was more strongly related to socioemotional outcomes of younger students in both proximal and distal domains. Younger students (with an average age of approximately 9 years) who participated in an arts program exhibited increases in the proximal domain of tolerance for others' perspectives, and in the more distal domains of growth mindset and academic goal orientation.

- Student Development Prior to Program Participation: Across ages, students with particularly high scores for certain distal domains of socioemotional development before participation realized a disproportionate benefit from arts education. For example, students who reported high levels of school engagement prior to participating in an arts program maintained these high levels of engagement. In contrast, students who had similar initial levels of engagement but who did not participate in an arts program demonstrated sharp decreases in school engagement. A similar pattern of findings was observed for academic self-efficacy, or students' perceptions of their capacity to succeed in school.

This is not to imply that program factors are unimportant. There was substantial variability in two factors: programs ranged in dosage from brief residencies featuring a few hours of instruction over the course of a month to year-long programs that met for over an hour most days of the week. The overall quality of staff instruction with respect to SEL also varied, with programs earning scores from slightly below 2 (on a scale of 1 to 5), indicating relatively infrequent occurrence of activities that should promote SEL, to above 4.

However, our analyses revealed no clear relation between program dosage and student outcomes, and only a loose coupling between the quality of instruction and outcomes. For example, the highest-rated program in our study achieved effects across multiple socioemotional outcomes, but so too did programs receiving

substantially lower ratings. This may mean that in a sample of very diverse students, student factors such as age are a more powerful influence on what a program can achieve.

III. CONCLUSION

Through this project the Foundation has made three contributions to the field of arts education. First, the project has contributed news tools to the field that can be used to continue to generate knowledge in the future. The observational tool developed for this study can form the core of a set of measures that will allow the Foundation to assess program practices designed to achieve outcomes across many domains, from aspects of socioemotional development not measured here to learning in the arts. Second, the project has contributed new knowledge to the field, yielding a more nuanced understanding of how – and for whom – arts education can foster socioemotional development. However, the context in which the Foundation has made these contributions is itself the most important contribution to the field. Our understanding of the effects of arts education on socioemotional outcomes is based largely on evidence collected from children who are more affluent and demographically homogenous than those whom the Foundation serves. By investing in the creation of knowledge among more diverse samples of students – as well as the measures necessary to generate this knowledge – the Foundation has begun to address the inequitable distribution of knowledge about how the arts may improve the lives of students.

THE SOCIOEMOTIONAL BENEFITS OF THE ARTS:
A New Mandate for Arts Education

I. CONTEXT

What is the mandate for arts education? Why does it matter if students receive an education in the arts?

One answer enshrined in public policy is that the arts are part of a "well-rounded education" (Every Student Succeeds Act, p. 807). This legislation asserts that just as we would consider a student's education incomplete without English Language Arts or science, we should regard an education without the arts as incomplete. This is, however, a dubious mandate. The place of English Language Arts is assured by the need for students to be literate. The place of science is cemented by the perceived future demand for technical skills. But the arts enjoy neither the advantage of an essential capacity nor the perception that they offer the benefit of imparting job skills (though they may, in fact, do so).

Instead, the cornerstone in the argument for arts education could be the more equitable distribution of educational opportunity. As evidence attesting to the central role of socioemotional skills in school and life success accrues (Farrington et al., 2012; Zimmerman, 2002), advocates have begun to argue that arts education may foster students' socioemotional development. If they are correct, one of the factors that may contribute to reduced school and life success among low-income students is their reduced access to arts education and concomitant restrictions on their opportunities to build socioemotional skills such as persistence and learning mindsets.

To begin this inquiry the William Penn Foundation (hereafter, the Foundation) engaged WolfBrown, an arts and arts education research firm, to collaborate with Johns Hopkins' Science of Learning Institute in an evaluation of a cohort of its 2015-2016 arts education grantees. This collaboration began with a review of literature on the capacity of arts education to foster socioemotional development, and then proceeded to conversations with grantees in which we formulated a set of socioemotional outcome domains that might be unique to the experience of arts education— as opposed to those that could result from other extra-curricular activities, such as athletics or science clubs— but that were *not* likely to be specific to a single arts discipline, such as theater. In the following report we present the results of this work.

A. The Broad Context for the Foundation's Work in Arts Education
The mission of the Foundation with respect to the arts is to foster a vibrant and creative city. To this end, the Foundation supports many of the Philadelphia area's major cultural institutions by providing flexible, unrestricted core support so that arts organizations can pursue their missions and make innovative, high-quality art work available to diverse audiences. The Foundation also provides funding for artistic projects that explore the use of public and civic spaces as venues for performances and public art that engages audiences in new ways. As part of this community-engaged portfolio, the Foundation's investments in arts education also serve multiple purposes with the over-arching goal of enabling more students in Philadelphia to engage with the arts and gain access to the region's rich cultural offerings. Depending on the goals of the Foundation's grantees, these may include exposing students to new art forms or art works,

providing an introduction to the skills required by different artistic disciplines, or supporting the development of students who are already engaged in the arts at a high level.

And yet across all these activities the Foundation and its grantees share commitment to harnessing the capacity of the arts to foster young people's sense of purpose, belonging, and possibility, and in so doing, to help them thrive. But in contemporary Philadelphia, thriving is far from assured. Among the nation's largest cities, Philadelphia has the highest poverty rate (26.3%), with a rate of deep poverty[2] (12.3%) that is twice the national average and affects approximately 60,000 children (Philadelphia Inquirer, 2016).

One feature of the developmental environment of young people placed 'at risk' by poverty has been the steady winnowing of experiences and opportunities that might foster socioemotional learning (SEL), which results in the acquisition of skills that are critically important for success in school and life, including the ability to manage behavior and make effective decisions, the maintenance of a positive self-concept, and the capacity to interact productively with others.

Socioemotional learning may occur in many contexts, including instruction in 'core' academic subjects such as reading or math, but the research we review below suggests that the arts offer a particularly fertile context in which SEL may occur. For example, the ebb and flow of overcoming successive challenges through the sustained exertion of effort is a part of learning to practice any art form, but it is also one way in which students may build perseverance. Similar, the experience of gradually mastering a particular artistic technique, if developed in a context of specific forms of positive feedback, may orient students' implicit theories towards the potential for growth.

As the schools that serve children in poverty have become increasingly focused on transmitting a discrete set of academic skills, the opportunity for SEL in arts education has also become less frequent, even to the point of absence (Parsad & Spiegelman, 2012). The reasons for this shift are many and varied, and include the proliferation of high-stakes testing, the competition for 'elective funds' (Beveridge, 2010), and the mechanisms through which schools are funded. But two consequences of this shift are clear: 1) the opportunities for SEL through the arts are distributed unevenly by affluence, and 2) given the known associations between socioemotional skills and school success, the uneven distribution of these opportunities further disenfranchises students already disadvantaged by their families' socioeconomic status.

B. Can Arts Education Foster Socioemotional Development? The Specific Context for the Evaluation

It is against this background that the Foundation has made investments in arts education for children attending school in Philadelphia. By devoting time and resources to this evaluation, the Foundation sought to address for the first time whether the programs of its grantees—delivered across a range of disciplines, dosages, and formats—create contexts for SEL.

To address this question, we first had to define socioemotional learning (SEL) within the context of the evaluation. As noted above, SEL is the process through which students acquire socioemotional skills, but this begs the question: what skills, or what specific domains of socioemotional development, should be the focus of the evaluation? From a practical perspective, assessing all the myriad skills associated with SEL was not possible. Doing so would require the administration of a prohibitively-long instrument, and not all

---

2 Deep poverty is defined as a family of three living on less than $10,000 a year.

socioemotional domains are best measured using measures for which students answer questions about themselves (i.e., self-report measures).[3]

Instead, we arrived at a set of socioemotional domains by employing two processes: a 'top-down' approach in which we reviewed the developmental and educational literature to identify how arts education might provide a context for SEL, and a convergent 'bottom-up' approach, in which we conferred with the leadership of the arts education programs in the cohort about the specific domains of socioemotional development that they believed their programs were most likely to influence. During these discussions we focused on a sub-set of socioemotional outcome domains that might be unique to the experience of arts education – as opposed to those that could result from extra-curricular activities more broadly, such as athletics – but that were *not* likely to be specific to a single arts discipline.

Based on this initial work, we identified three broad research questions:

1. Can arts education programs foster students' socioemotional development in "proximal" domains?
Domains of socioemotional development that are proximal to the arts include interest in the arts, tolerance for others' perspectives, and cultural awareness. For example, relatively intensive programs of drama education featuring instruction over the course of multiple weeks have been associated with significant increases in tolerance among both elementary- (Gourgey, 1985) and high-school students (Beales & Zemel, 1990). However, even visual arts education programs of very limited dosage might influence these domains: students who participated in a field trip to an art museum exhibited higher levels of interest in the arts, tolerance, and historical empathy than their peers (Greene, Kisida, & Bowen, 2014).

2. Can arts education programs foster socioemotional development in more "distal" domains?
Some arts education program leaders also proposed that their programs could influence more distal domains of socioemotional development, including perseverance, school engagement, growth mindset, academic goal orientation, academic self-concept, and academic self-efficacy. With varying degrees of confidence, program leaders asserted that their programs could influence these domains in a general (i.e., *not* arts-specific, as in the case of perseverance) or academic context. While it is often difficult to achieve this kind of 'far transfer' (Barnett & Ceci, 2002), it was important to investigate this possibility. To this end, we selected a set of distal socioemotional domains that all programs, regardless of discipline or format, could potentially foster:

- **Perseverance** refers to one's ability to exert sustained effort in the face of obstacles or in the course of difficult or lengthy tasks (Peterson & Seligman, 2004). Although perseverance is thought to be a relatively stable personality trait, Winner and Hetland (2008) have argued that an education in the arts may provide opportunities to engage in challenging but nevertheless rewarding tasks, and that repeated experiences of this type may foster perseverance in the context of the arts over time (Winner & Hetland, 2008). However, to date only a small number of studies have yielded empirical evidence to support the argument that arts instruction may lead to increased perseverance in more general contexts.

  Participation in instruction featuring both visual arts and music activities over the course of three years was associated with more rapid growth in teacher reports of persistence among

---

3 For example, while students' responses on self-report measures may provide insight into their perception of their social and interpersonal skills, a more valid measure of their *actual* skills might involve behavioral observation.

Scandanavian elementary-school children (Metsäpelto & Pulkkinen, 2012), while South African adolescents from low-income families reported higher levels of perseverance following a program of musical instruction (Devroop, 2012). Similarly, Scott (1992) used behavioral measures to examine perseverance among preschoolers (ages 3 to 5 years) who had taken either private or group Suzuki-based violin instruction for at least five months. Students enrolled in violin instruction, whether privately or in groups, scored higher on measures of perseverance than their control-group peers. In summary, while there is some support in the literature linking arts education to perseverance, that literature is drawn from a small number of studies, most of which feature music education.

- **School Engagement** is a multi-faceted construct that includes students' participation in extra-curricular activities, feelings about school, and investment and valuation of learning (Fredericks, Blumenfeld, & Paris, 2004). Multiple researchers have proposed that school engagement influences success in school (Connell, 1990; Finn & Voelkl, 1993), and indeed higher levels of school engagement have been found to predict lower rates of school dropout, better grades, and higher standardized test scores (Fredricks et al., 2004; Marks, 2000).

  A small number of correlational studies (Center for Music Research, 1990; Baum & Owen, 1997; Csikszentmihalyi & Schneider, 2000) have demonstrated that students exhibit higher levels of investment and engagement in the classroom while participating in arts classes, as opposed to classes on academic subjects, but this tells us little about how the arts contribute to students' overall engagement in school. It may be that students exhibit increased engagement while participating in arts classes, but that this fails to transfer to the more general school context.

  Additionally, both correlational (Heath, 1998) and experimental (Walker, Tabone, & Weltsek, 2011)[4] studies indicate that arts instruction predicts fewer absences from school among middle- and high-school students, even among high-school students deemed at risk for dropping out of school (Center for Music Research, 1990; Horn, 1992). Studies using direct assessments of students' self-reported school engagement have yielded similar results. For example, a correlational study of high school students revealed that after controlling for relevant covariates students who participated in the arts reported higher levels of school engagement than their peers (Elpus, 2013).

  Similarly, two large, quasi-experimental studies have also supported the potential for arts education to foster school engagement. The Drama Improves Lisbon Key Competencies in Education (DICE Consortium, 2010) study was a large (over 5,000 students), quasi-experimental study that examined the effects of drama education on a number of student outcomes, and which demonstrated that students who participated in drama education reported more positive feelings about school than their peers. In another large study, Smithrim and Upitis (2005) investigated the effects of a whole-school model of arts-integration on students' school engagement. Students in grades 1 through 6 who were attending one of 55 treatment or 35 comparison schools were asked to rate their attitudes about school at the beginning and end of the school year. Girls in the sixth grade who attended treatment schools reported significantly-higher levels of school engagement than their comparison-group peers. Thus there is both correlational and quasi-experimental support for a link between arts education across disciplines and higher levels of school engagement.

---

4 Unlike the other studies reported in this paragraph, Walker and colleagues (2011) examined the effects of a theater-integrated ELA curriculum, rather than a general program of arts education.

- **Growth mindset** refers to the extent to which an individual sees their intelligence as mutable, as opposed to fixed (Dweck, 2000). According to a motivational model of achievement, those who see their intelligence as open to change will be more inclined to seek out challenges and thereby maximize their potential, while those who believe their intelligence is fixed will be more likely to avoid situations in which their limitations may be exposed. This approach entails avoiding challenges, limiting one's opportunity to improve and resulting in the fixed mindset becoming a self-fulfilling prophecy. However, research has demonstrated that students' implicit theories of intelligence are open to change, and that embracing a growth mindset predicts better academic performance (Aronson, Fried, & Good, 2002). It is therefore possible that an education in the arts may alter students' implicit theories of intelligence, as they experience changes in their own artistic skill or understanding, supported by feedback that fosters a growth mindset. However, at present no study has assessed this possibility.

- **Goal orientation** is a very broad concept that describes why and how people are compelled to achieve an objective (Anderman & Maehr, 1994; Kaplan & Maehr, 2006). Generally, three types of goal orientations are recognized: performance-approach, performance-avoidance, and mastery goals. Of these, an orientation towards mastery goals – a desire to develop competence, rather than to appear intelligent (i.e., performance approach) or avoid appearing foolish (i.e., performance-avoidance) – have been most consistently associated with behaviors that support learning and academic achievement (Ames, 1992). However, only one study has examined change in **mastery goal orientation** as a function of arts education: French-Canadian middle-school students who participated in a French-immersion, drama-based course of instruction exhibited higher levels of growth in one self-report measure of motivation than students receiving a more conventional (but still French-immersion) course of instruction on the same topic (Bournot-Trites, Belliveau, Spiliotopoulos, & Séror, 2007). Thus the literature linking arts education to mastery goal orientation is quite limited.

- Another aspect of achievement motivation is **self-concept**, which is similar to self-esteem, but is generally considered to be domain-specific: thus a bright but introverted student may have a positive academic self-concept, thinking highly of him- or herself with respect to school performance, but a negative social self-concept, doubting his or her ability to interact effectively with peers. There is a relatively large literature linking participation in arts programs to improved general self-concept (or self-esteem), though there are single-group (Kander, 2009; Kaufman et al., 2014; Mason & Chuang, 2001), correlational (Respress & Lufti, 2006; Wright, 2006), and experimental (Costa-Giomi, 2004; Minton, 2001) studies, and some of these studies focused on students from low-income families or who were otherwise deemed at risk for failure in school (Costa-Giomi, 2004; Mason & Chuang, 2001; Respress & Lufti, 2006).

  However, there are also multiple studies that link arts education to the more focused concept of academic self-concept. Correlational studies of American high-school students have revealed an association between courses of both multi-arts (Buton, Horowitz, & Abeles, 2000; Catterall, 1998) and dance (Carter, 2005) instruction and higher levels of academic self-concept. A similar association was reported between the duration of musical instruction and academic self-concept with slightly younger children (ages 12 to 14 years), and this association was robust even after controlling for other factors such as family socioeconomic status (Degé, Wehrum, Stark, & Schwarzer, 2009).

Studies with groups of low-income middle school students in the U.S. (Shin, 2011) and Australia (Russell-Bowie, 2013) have revealed an association between academic self-concept and participation in programs of music and multi-arts instruction, respectively. McPherson and colleagues extended these findings to a large sample (over 2,500) of middle- and high-school students studying music (McPherson, Osborne, Barrett, Davidson, & Faulkner, 2015).

However, to date few studies have employed quasi-experimental or experimental methods to investigate whether participation in the arts may *cause* higher levels of academic self-concept. Lee (2007), for example, demonstrated that dance instruction was related to higher levels of academic self-concept in a sample of school-aged students in South Korea. Rickard and colleagues reported similar results for a group of young children (age 8) participating in a school-based music education program (Rickard et al., 2013).

In summary, there is both correlational and quasi-experimental evidence for a link between arts education and academic self-concept; however, unlike school engagement, the available quasi-experimental evidence is drawn from smaller studies examining the relation between instruction in one arts discipline and academic self-concept. It is therefore more difficult to generalize the results of these studies to instruction across disciplines offered to students from a wide array of backgrounds.

- **Self-Efficacy**, like self-concept, is generally considered to be domain specific. Thus a student who excels at academics but not visual art may report feeling quite efficacious with regard to school but far less so with regard to painting. Like perseverance and achievement motivation, higher levels of self-efficacy in a particular domain predict better performance in that domain, above and beyond actual capacity (Bandura, 1986).

  In a quasi-experimental study, Lee (2006) demonstrated that students attending elementary school in South Korea reported higher levels of academic self-efficacy than their peers following a program of dance instruction. Similar results were reported by Rapp-Paglicci and colleagues, who found that participation in a course of multi-arts instruction was associated with increased academic self-efficacy among a sample of adjudicated youth (Rapp-Paglicci, Stewart, & Rowe, 2011).

  In two studies Catterall and his colleagues founds that participation in the arts was associated with students' general or overall self-efficacy. In one study, Catterall and Peppler (2007) examined self-efficacy among low-income third-grade students who participated in a program of high-quality visual arts instruction. The authors found a significantly larger proportion of students in the treatment group (approximately half) made significant gains in self-efficacy, relative to the comparison group. In a second study Catterall (2007) reported that low-income students who participated in a theater education program evidenced significantly larger gains in general self-efficacy than a comparison group of their peers recruited from the same schools. The link between arts education and academic self-efficacy is therefore supported by two studies, both of which involved participants who were not similar to the students served by program grantees (in that one study was conducted in South Korea and the other with adjudicated youth). Studies with students more similar to those served by grantees support a link between arts education and general, rather than academic, self-efficacy.

To some extent the opinions of the leaders of the arts education programs about the capacity of their programs to foster socioemotional development in particular domains varied as a function of their awareness of them. For example, most leaders felt confident that their programs could foster school engagement. However, for other domains about which leaders were equally aware—perseverance, for example—the views of leaders were more diverse. Leaders of organizations that offered shorter, less-intense programs were less confident that their programs would foster perseverance. In contrast, leaders of programs that featured sustained instruction in an artistic discipline were more confident that perseverance might be an attainable outcome.

3. Can arts education programs foster socioemotional development in domains of artistic self-awareness such as artistic goal orientation, self-concept, or self-efficacy?

Ironically, less is known about whether arts education programs might influence artistic self-awareness, or the development of *artistic* goal orientation, self-concept, or self-efficacy. For example, might a program of music education influence how students see themselves in the context of music or how efficacious they feel regarding their ability to learn about music? Previous research has found that children at risk for educational failure (Shields, 2001) and male students who were incarcerated (Kennedy, 1998) who received musical instruction reported significant increases in musical self-concept, and that teacher-reports of higher musical aptitude predicted more rapid growth in students' musical self-efficacy over time (Zelenak, 2015). But beyond these studies only one of which we are aware reported an association between instruction in an artistic discipline other than music and artistic self-concept (visual arts; King, 1983).

Given this, we included measures of students' *artistic* goal orientation, self-concept, and self-efficacy among our measures, aligned in each case to the artistic discipline(s) in which instruction was offered. However, we administered these measures only to older students (i.e., those in high school), as we judged that these students would be most capable of distinguishing between academic and artistic contexts.

C. Goals of the Evaluation

The goals of our work with the Foundation were two-fold. First, we sought to answer the three guiding questions for the evaluation outlined above. To this end, data were collected from over 1,000 students attending a program offered by a Foundation grantee during the 2015-16 academic year. Prior to and following their participation in the program, students and their primary academic teachers were asked to complete measures of the socioemotional domains outlined above. For additional detail on these measures, please see the report on the pilot study conducted in the spring of 2015, attached as an appendix to this report; for additional detail on the sample, please see the technical appendix.

Second, we endeavored to provide the Foundation with a set of tools that it could use in the future to better understand the impact of its grantees working in the area of arts education. We therefore developed two sets of measures that are described in more detail below:

- A battery of student- and teacher-report measures that, as we describe below, were demonstrated to produce reliable data on socioemotional domains for students across a wide range of ages and backgrounds.

- An observational measure of the quality of teaching artists' practices and students' responses with regard to SEL. This measure was found to be internally consistent, and could be used reliably by multiple trained observers to identify promising practices. As such, it holds promise as a template and basis on which the Foundation could build a more expansive measure with which to gauge the quality of instruction along multiple dimensions.

II. FINDINGS

The research questions outlined above were addressed in two phases. First, we examined whether there were overall or omnibus effects of arts education on students' socioemotional development in proximal and distal domains. However, given the tremendous diversity of programs and the students they served, we did not expect to observe many of these omnibus effects. Rather, we expected that program effects would be contingent upon program factors—such as the intensity of dosage of instruction—and student factors, such as age. Therefore the second portion of our analyses focused on the role that program and student factors played in influencing the effects of arts education on students' socioemotional development.

A. Can Programs Foster Socioemotional Development for All Students?

Our first set of analyses examined whether there were omnibus effects of participation in arts education on post-program measures of both proximal and distal socioemotional domains.[5] Omnibus effects are effects that apply to all students, regardless of student factors such as age or program factors such as dosage.

The results of these analyses indicated that across all students there were effects for interest in the arts and growth mindset that approached statistical significance.[6] In both cases, the effect of participation was positive: students participating in arts programs exhibited higher post-program levels of interest in the arts and growth mindset after controlling for pre-program levels of these constructs. In both cases, the size of the effect for participation was modest.[7]

Figure 1 displays the average level of interest in the arts reported by students before and after the program, while Figure 2 presents the same for growth mindset. All measures used the same scale, ranging from 1 for strongly-disagree to 5 for strongly agree.

Figure 1. Interest in the Arts[8]
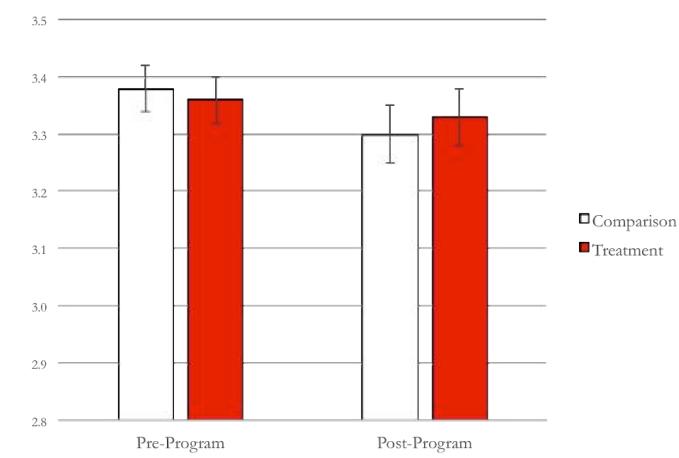
While both groups exhibited a decrease in scores from pre- to post-program, the magnitude of the decrease was smaller for students in the treatment group (red bars), and therefore the effect of participation was positive.

---

5 This question was examined in a series of multi-level models. Given that we were primarily interested in the discrete effects of current participation on socioemotional development all models accounted for students' levels of socioemotional development in a particular domain prior to program participation. This approach is often referred to as estimating residualized change. All models accounted for the nested structure of the data (in which students were grouped within classrooms which were in turn grouped within schools) as well as differences in groups by student age, gender, race/ethnicity, and prior exposure to the arts in school. While all models controlled for these student factors, models also controlled for students prior exposure to the arts out of school and their current arts instruction when those factors were related to a particular domain of socioemotional development. For additional detail on the models, see the technical appendix.
6 "Approached statistical significance" indicates that while there is a reasonable likelihood that the results we observed were not simply due to chance, that likelihood does not reach the 5% threshold commonly used to classify a result as "statistically significant."
7 According to the guidelines established by Cohen (1988).
8 Note that here and throughout error bars correspond to two standard errors about the mean.

Figure 2. Growth Mindset
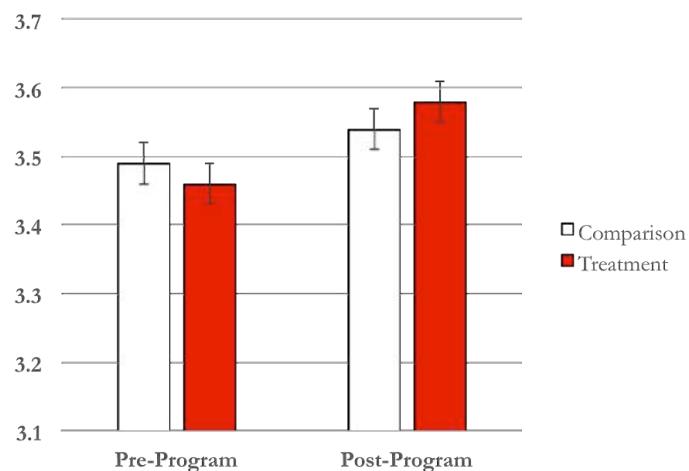
For growth mindset, both groups exhibited an increase, but the increase was larger among treatment-group students than their comparison-group peers.

As noted above, students enrolled in programs serving high school students were asked to report their goal orientation, self-efficacy, and self-concept in both school and in the context of the arts (and specifically the discipline(s) in which their program offered instruction). A nearly-significant effect was observed for artistic goal orientation among these students. In contrast to the effect sizes for interest in the arts and growth mindset, the size of this effect was moderate to large. Figure 3 presents student-reported artistic goal orientation pre- and post-program.

Figure 3. Artistic Goal Orientation Among High School Students
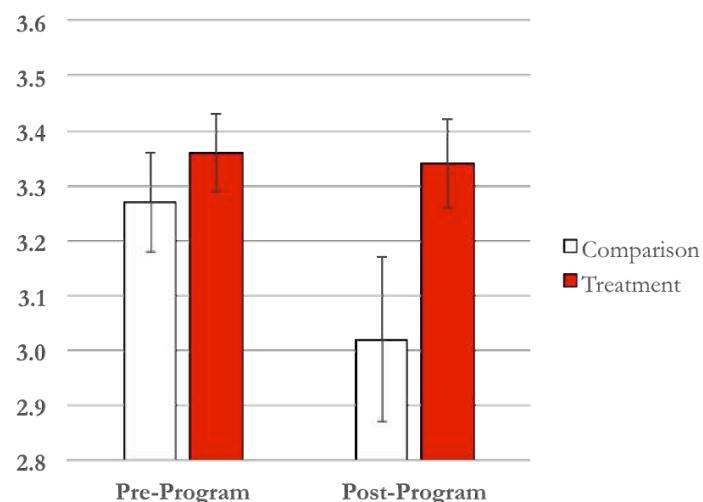
In this case students who were not enrolled in a program exhibited a decline in artistic goal orientation scores, while those enrolled in the program maintained stable levels of goal orientation.

B. Why Were Additional Omnibus Effects Not Observed? Understanding the Diversity of Programs

In reviewing the results above, it is reasonable to ask why additional or larger omnibus effects were not observed. When considering this question, it is important to bear in mind that an omnibus effect assumes that an effect will be observed for all students across all programs. When evaluating a single program offered to a relatively homogeneous group of students, this assumption may be reasonable. But in the context of this project, where there is great diversity in both programs and the students those programs serve, the expectation of main effects may be unrealistic. It is more likely that program effects will be contingent on the characteristics of both programs and students.

The programs included in the cohort for our study varied by many factors, including their history, the number of children they served, and the length of their partnership with a given school. We chose to focus on three factors that differentiated programs from one another: the discipline (or disciplines) in which arts instruction was offered, the dosage or intensity of the treatment, and the quality of SEL practices and student responses to them.

Artistic discipline(s) and dosage were relatively easy to assess using records provided by the Foundation and observations of programming as implemented. Table 1 presents the distribution of programs by these factors.

Table 1. Distribution of Programs by Discipline & Dosage

Program	Discipline(s)	Min./Session	Sessions/Week	Weeks/Year	Hours/Year	Classification
Program A	Music Visual Arts	90	3-4	36	162-216	High
Program B	Music	45	1	6	4.5	Low
Program C	Dance	45	1	16	12	Medium
Program D	Visual Arts	90	2-3	6	18-27	Medium
Program E	Dance	45	1-2	36	27-54	High
Program F	Theater	45	2-3	8	18	Medium
Program G	Visual Arts	45	1-2	6	4.5 – 9	Low
Program H	Dance	45	1	36	27	High
Program I	Visual Arts	60	1	16	16	Medium
Program J	Dance Theater	45	1	16	12	Medium
Program K	Theater	35	1-2	5	3 – 6	Low
Program L	Theater	45	1-2	5	4 – 7.5	Low
Program M	Theater	45	1	13	10	Medium

Low dosage: < 9 hours instruction per year
Medium dosage: 9 to 27 hours instruction
High dosage: > 27 hours instruction

Assessing the quality of instruction as implemented with regard to socioemotional learning required the development of a structured observation protocol that drew on existing models (e.g., the Youth Programming Quality Assessment, or YPQA; Smith, 2012) as well as research on the practices that may

foster young people's socioemotional development (e.g., Benson et al., 2010; Catalano et al., 2004; Roth et al., 1998). The adapted measure was organized into the following broad dimensions:

- **Context:** The physical and interpersonal environment in which the teaching and learning occurs.

- **Challenge:** The degree to which the structure, content, and processes of working offer young people the engagement, high expectations, and scaffolding required to try new materials and skills, make meaning, identify and develop their own ideas.

- **Belonging:** A young person's sense that they are a welcome and acknowledged member of the classroom or school community, regardless of who they are, or what their current level of skill is. Correspondingly, this refers to the ways in which youth, as well as teachers acknowledge others.

- **Relevance:** The extent to which assignments, materials, ways of working, and models and examples acknowledge who young people are, and the extent to which young people themselves forge connections to their own lives and experiences.

- **Self-efficacy:** The observable behaviors that index teachers' and students' understanding of and commitment to strategies for setting and reaching goals in specific situations (e.g. drawing, writing a personal narrative, singing, etc.).

- **Growth Mindset:** The extent to which teachers and students emphasize growth mindsets, in which capacity can be expanded with effort, persistence, revision, and risk taking, over fixed mindsets, in which capacity is immutable or innate.

Each of these broad dimensions was further divided into a series of subscales, listed in Table 2 below.

Table 2. Scales and Subscales for the Observational Measure

Context	Challenge	Belonging
- Space	- Explanation	- Tone
- Materials	- Appropriate Time	- Interaction*
- Teacher Partnership	- Higher-Order Thinking*	- Collaboration*
	- Larger Projects*	
Relevance	**Self-Efficacy**	**Growth Mindset**
- Exchange*	- Skill Focus*	- Challenge*
- Connection*	- Goal Setting*	- Persistence*
- Choice/Generativity*	- Improvement*	- Growth-Oriented Response*
	- Leadership*	- Mindset Reflection
	- Critical Assessment*	

Note: Asterisks indicate subscales that were rated independently for both staff and youth.

All programs were observed by raters who had been trained to an acceptable degree of reliability; as such, we had a high level of confidence that the scores assigned to a program by one rater would be very similar to the scores assigned by another rater. Shorter programs were observed once at their midpoint, while

longer programs were observed twice (half-way and two-thirds through their duration) and the average of the two observations was used as that program's score. For more details about the collection of data using the observation measure of quality with respect to SEL, see the technical appendix.

Overall measures of staff and youth quality with respect to SEL were calculated for each program. Given that staff and student scores were closely correlated, staff measures are reported for all programs in Figure 4 below:

Figure 4. Quality of Instruction by Staff with Respect to SEL[9]
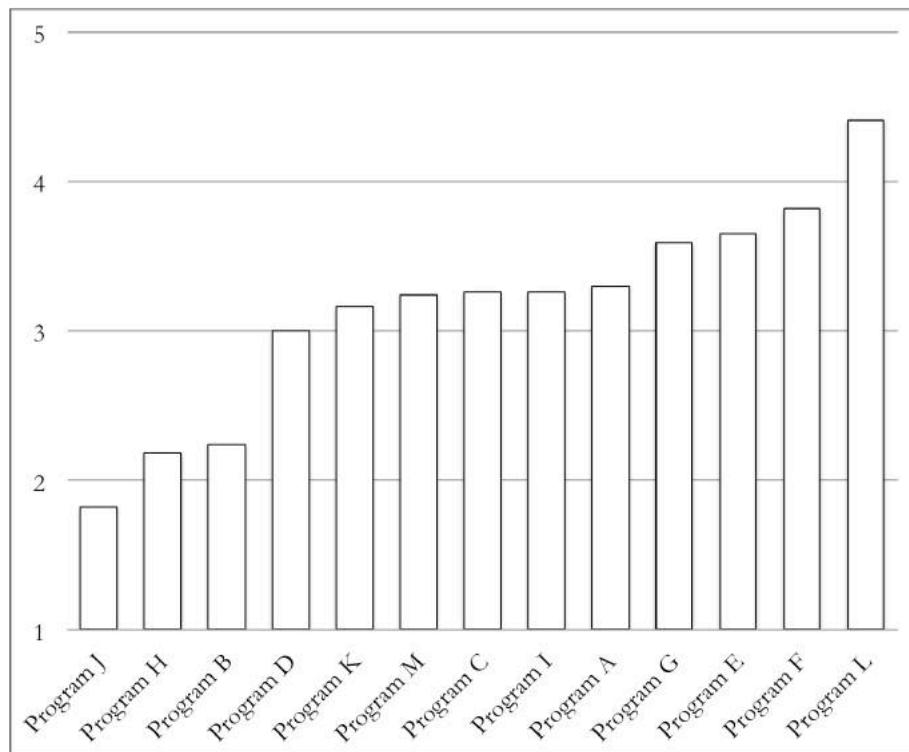
Figure 4 reveals two things about the Foundation's portfolio of grantees. First, while no programs earned the lowest or highest possible score on the measure, we observed nearly the full range of possible values on the measure. Thus there were a small number of programs (e.g., Program J) for which the overall score was approximately a 2 out of 5 (indicating a relatively low level of staff behaviors that would foster SEL), while there was one program (Program L) with an overall score that exceeded 4 (indicating a relatively high level of these behaviors).

Second, there was a cluster of six programs that scored at approximately the midpoint of the scale. While this is reassuring in that most programs therefore meet or exceed this benchmark for quality of instruction with respect to SEL, it also makes linking variability in quality of instruction to students' SEL more difficult, as there is less variability than might otherwise have been observed.

9 Although overall staff and youth behavior scores were calculated, there was a very close correspondence between scores assigned for each program. Therefore we report overall staff scores in Figure 4 and use them throughout.

What Figure 4 does not reveal is that at the level of the scales and subscales that comprised the measure, even programs that were assigned low overall ratings of quality nevertheless scored relatively-well on some scales of instructional quality. For example, despite its low overall quality rating, staff at Program J received similar scale scores on fostering student belonging as programs that earned average overall quality ratings. Conversely, even programs that earned high overall scores (e.g., Program L) scored below the mean on instructional context.

C. What Is Revealed by Examining the Effects of Each Program?

With this information regarding the quality of instruction with respect to SEL providing valuable context, we can now consider the results observed within each program. These results were obtained using the same analytic procedures as those employed in the cross-program analyses, where post-program measures of each domain (interest in the arts, tolerance, etc.) were estimated after accounting for students' scores on these domains prior to program enrollment.

These analyses did not yield a large number of significant main effects using the standard threshold to establish statistical significance. This is not surprising, given that the relatively small number of students served by each program and the need to account for relevant covariates reduces the likelihood that significant effects will be observed. However, if we focus on effects that exceeded the benchmark for "educational significance" (d = .25, Tallmedge, 1977) we can see that many programs achieved multiple of these 'sub-threshold' main effects (see Figure 5).

Figure 5. Summary of Main Effects by Program[10]

Program	PROXIMAL — Interest in the Arts	PROXIMAL — Tolerance for Others	PROXIMAL — Cultural Awareness	DISTAL — Perseverance	DISTAL — Growth Mindset	DISTAL — School Engagement	DISTAL — Academic Goal Orientation	DISTAL — Academic Self Efficacy	DISTAL — Academic Self-Concept	TRANSFER — Artistic Goal Orientation	TRANSFER — Artistic Self Efficacy	TRANSFER — Artistic Self-Concept	DISTAL (Teacher) — School Engagement	DISTAL (Teacher) — Perseverance
Program A	√	√				√								
Program B					√		√						*	√
Program C					√									
Program E													√	√
Program F[T]								*						
Program H	√	√		√	√			*						
Program I														
Program J		√		√	√		√		√				√	√
Program K														
Program L			√							√T	√T			*
Program M			√				√							√

√ indicates a sub-threshold main effect, while * indicates a significant main effect.

[T] indicates that a program or domain employed a single-group (treatment only) design

---

10 Two programs included in Figure 4 – those offered by Programs D and G – are not included in Figure 5, and indeed the data collected from these programs were excluded from the analyses of the combined program data due to concerns about their integrity. For additional details, please consult the technical appendix.

Three points emerge from these findings:

- **Sensitive Domains:** Looking down the columns we note that effects were more likely to be observed for some domains of socioemotional development than others. Interestingly, effects on the distal domains of growth mindset and teacher-rated perseverance were most likely to be observed, rather than the more proximal domains interest in the arts or artistic goal orientation, self-efficacy, or self-concxept.

- **Program Profiles:** Looking across rows it is clear that a number of programs achieved multiple sub-threshold main effects. For example, students in Program B realized sub-threshold main effects for growth mindset, academic goal orientation and teacher-rated perseverance as well as a significant effect for teacher-rated school engagement. A similar number of effects were observed for Program H and Program J. This suggests that programs may have particular areas of strength and weakness. Building communities of practice and promoting intentional design to exploit strengths and redress weaknesses could enhances programs' effects on students' socioemotional development.

- **Loose Coupling of Program Practice and Outcomes:** Comparing the prevalence of significant and sub-threshold main effects to observational ratings of program quality reveals a 'loose coupling' between the observed quality and the presence of these effects. Program L, for example, achieved sub-threshold main effects for cultural awareness, artistic self-efficacy, artistic self-concept, and a significant effect for teacher-rated perseverance. This relatively large number of effects is consistent with the high overall rating of quality with respect to SEL assigned to Program L. But other programs that were assigned a much lower rating, such as Programs B and J, also achieved multiple effects. This suggests that the observational measure should be refined to consider other factors such as classroom and school climate.

D. Moving Beyond Program Factors to Consider the Diversity of Students

This pattern of results strongly suggests that other factors beyond quality are at work in the relation between program enrollment and socioemotional outcomes. Dosage does not appear to be the key factor: comparing the pattern of effects in Figure 5 to the dosage information reported in Table 1 reveals no discernable relationship between these factors and the likelihood of observing effects.[11] But other factors such as student characteristics could be working either alone or in concert with these program factors to produce these effects.

When completing their surveys students were asked to indicate their date of birth, gender, race/ethnicity, and whether they had ever had in- or out-of-school instruction in the artistic discipline(s) offered by the program in which they would be enrolled. Students' primary or homeroom teachers were also asked to provide information regarding the arts instruction their students would receive in the coming year.

Using this information we can say that the 'average' student in our study was approximately 12 years old at the time she or he completed their pre-program survey, was slightly more likely to be a girl than a boy, and had a 3 in 5 chance of being either African American or Latino/Hispanic. The average student had a 3 in 5 chance of having had in-school arts instruction in the discipline(s) offered by the program and a 2 in 5

---

11 A series of multilevel models that introduced dosage and observed quality with respect to socioemotional development confirmed these impressions: in no case did either of these program factors interact with participation.

chance of having had out-of-school instruction in the same, and was most likely to have in-school instruction in visual arts and music 1 or 2 times a week for approximately half the 2015-16 school year (for additional detail on the composition of the sample, see the technical appendix).

Of course, the 'average' student is an abstraction. In practice there was a great deal of diversity among our students, who ranged in age from 8 to 18 years, and many of whom had considerably less arts instruction during the 2015-16 academic year than that reported for the average student. We therefore examined whether the effects of enrollment in an arts education program might vary as a function of these student factors.

E. Do Program Effects Vary by Student Factors?
Our analyses revealed that the effects of enrollment on some domains of socioemotional development were contingent upon students' age and pre-program levels of socioemotional development (for additional detail on the results of these models, see the technical appendix).

Younger students exhibited higher levels of growth mindset post-program than their comparison-group peers. For students enrolled in these arts education programs, growth mindset levels increased over time; for students who were not enrolled, these levels decreased. By the time of post-program measurement the magnitude in this difference corresponds to a large effect size. These diverging trajectories are displayed in Figure 6 (there and throughout, solid red lines indicate the treatment group of students who were enrolled in the arts education program).

**Figure 6**. Growth Mindset Among Younger Students
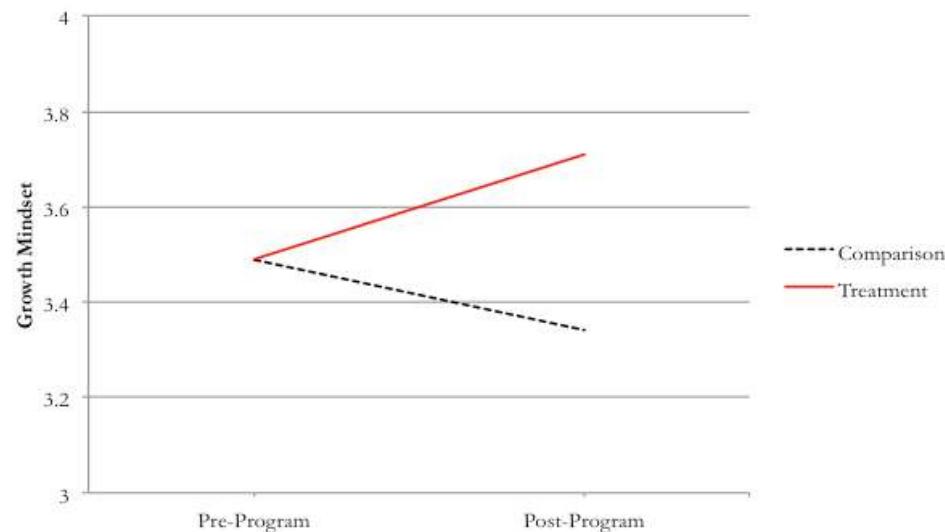
A similar pattern of results was observed for students' tolerance for others' perspectives and their academic goal orientation, though here the effects approached, rather than achieved, statistical significance (see Figure 7).

Figure 7. Tolerance and Academic Goal Orientation Among Younger Students[12]
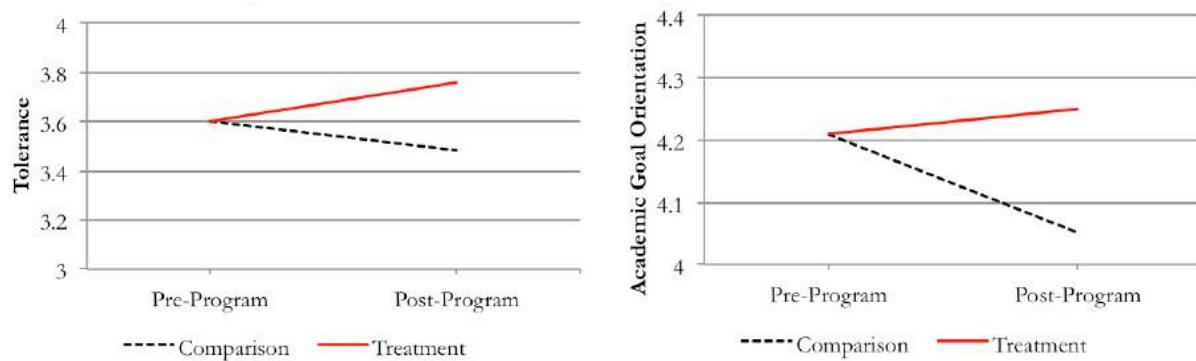
It was also the case that students who exhibited particularly high scores in certain domains prior to program implementation realized a disproportionate benefit from program enrollment. Figure 8 presents student-reported school engagement for participants with high pre-program scores. Regardless of whether students are enrolled in a program offered by the Foundation's grantees they exhibit lower school engagement scores over time. However, the decline in scores for students in the comparison group is more precipitous than that observed for their treatment-group peers. By the time of post-program measurement the difference in school engagement scores has reached a size corresponding to a moderate effect.

Figure 8. School Engagement Among Students with High Pre-Program Scores
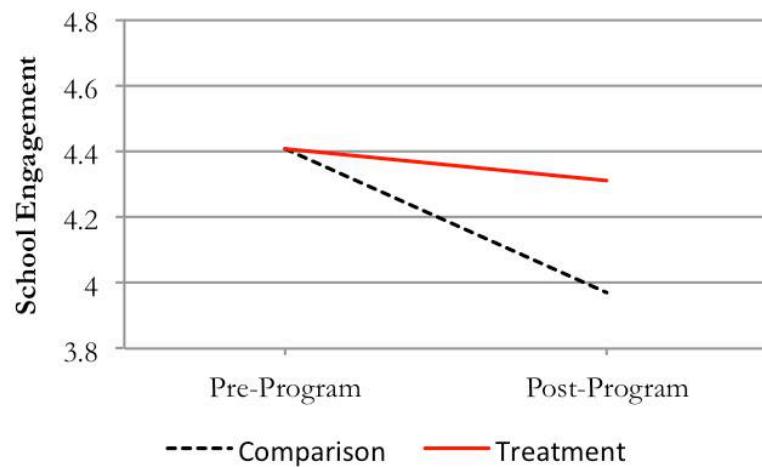
A very similar pattern of results was observed for academic self-efficacy (see Figure 9). Students with high pre-program scores who participated in a program exhibited a decline in scores over time, but this decline was slight relative to that observed among students not enrolled in these programs. The magnitude of the difference between groups at the post-program measurement corresponded to a moderate effect.

---

12 Note that although the two panels of Figure 7 are presented side by side, they do not share a common scale along the y-axis.

Figure 9. Academic Self-Efficacy Among Students with High Pre-Program Scores
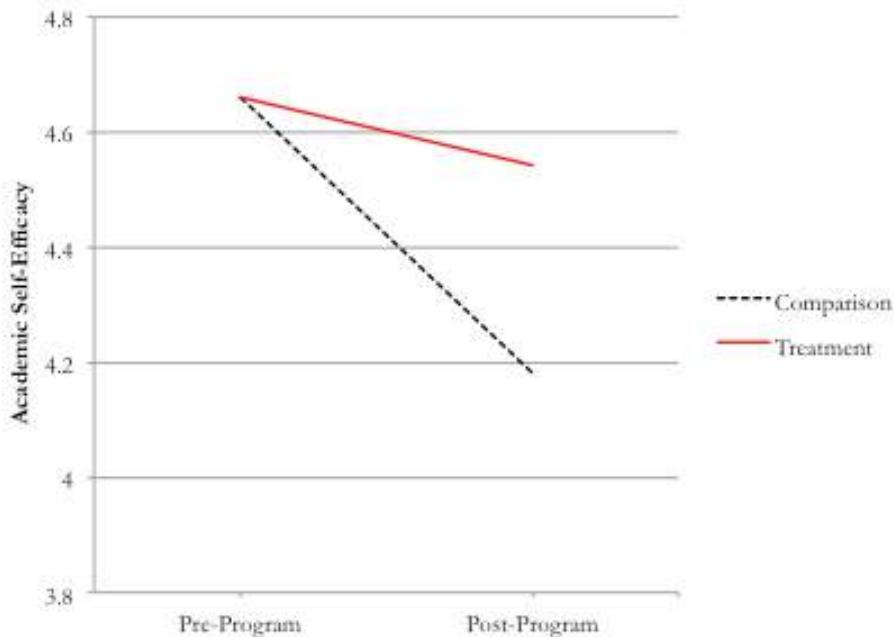
F. Is There a Relation Between Prior or Current Arts Instruction and Socioemotional Development?

As noted above, the purpose of these analyses was to examine whether participation in programs offered by a cohort of the Foundation's arts education grantees was associated with differences in socioemotional outcomes. In particular, we were interested in whether there was a *unique* association between participation and socioemotional outcomes after accounting for other factors, such as gender and school. One set of factors we controlled for was students' current and prior exposure to the arts in school. In doing so, we observed that there were positive relationships between current and prior arts exposure and a number of socioemotional domains. The question then became: how strong are these relationships, and to what extent are they unique?

To answer this question we estimated a series of models that predicted select socioemotional domains as a function of prior or current arts exposure after accounting for students' age, gender, race/ethnicity, classroom, and school. The results of these models indicated that there were unique relationships between prior in-school arts exposure and students' interest in the arts, awareness of other cultures, academic self-efficacy, academic self-concept, and engagement, with students who had prior exposure to the arts in school reporting higher scores in each of these domains. In each case, the size of the relation between prior exposure to the arts and the socioemotional domain in question was modest.

A similar set of relationships was observed between current exposure to the arts and both academic self-concept and school engagement. In both cases, higher levels of current in-school arts instruction was associated with higher levels of socioemotional outcomes in these domains. While the relationship between current arts instruction and academic self-concept was modest, the relationship between instruction and school engagement was moderate. For additional details regarding these models, please see the technical appendix.

III. CONCLUSION

Through this project the Foundation has made two contributions to the field of arts education.

First, the project has contributed new knowledge to the field. While the fact that the Foundation achieves a more equitable distribution of educational opportunity was not in doubt, prior to this project it was not clear whether these opportunities could foster students' socioemotional development. The results presented here indicate that they can, but that effects for all students are most likely to be observed in domains proximal to the experience of arts education, such as interest in the arts. Effects in more distal domains are possible, but they are most likely to be observed among sub-sets of students, including younger students and students exhibiting high levels of socioemotional development prior to program participation.

Second, the project has contributed new tools to the field that can be used to continue to generate knowledge in the future. The measures administered to students and teachers in this study produced reliable data across an array of socioemotional domains, and can be expected to do so again in the future. But a more substantial contribution is represented by the observational measure. The expanded version of the measure summarized in this report could form the core of a set of strategies that will allow the Foundation to assess program practices designed to achieve outcomes across many domains, from aspects of socioemotional development not measured here to learning in the arts.

Finally, the context in which the Foundation has made these contributions is itself the most important contribution to the field. As noted above, our understanding of the effects of arts education on socioemotional development is based largely on evidence collected from children who are more affluent and demographically-homogenous than those whom the Foundation serves. With some exceptions, the measures used to assess socioemotional development were also created with less diverse groups of students. By investing in the creation of knowledge among more diverse samples of students – as well as the measures necessary to generate this knowledge – the Foundation has begun to address the inequitable distribution of knowledge about how the arts may improve the lives of students.

WORKS CITED

Ames, C. (1992). Classrooms: Goals, structures, and student motivation. *Journal of Educational Psychology, 84,* 261-271.

Aronson, J., Fried, C. B., & Good, C. (2002). Reducing the effects of stereotype threat on African American college students by shaping their theories of intelligence. *Journal of Experimental Social Psychology, 38,* 113-125.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory.* Englewood Cliffs, NJ: Prentice-Hall.

Baum, S., Oreck, B., McCartney, H. (1999). Artistic talent development for urban youth: The promise and the challenge. In E. Fiske (Ed.). *Champions of change: The impact of the arts on learning.* Washington, DC: Arts Education Partnership & The President's Committee on the Arts and Humanities.

Beales, J. N., & Zemel, B. (1990). The effects of high school drama on social maturity. *School Counselor, 38,* 46-51.

Benson, P. L., Scales, P. C., & Syvertsen, A. K. (2010). The contribution of the developmental assets framework to positive youth development theory and practice. *Advances in Child Development and Behavior, 41,* 197-230.

Beveridge, T. (2010). No Child Left Behind and fine arts classes. *Arts Education Policy Review, 111,* 4-7.

Bournot-Trites, M., Belliveau, G., Spiliotopoulos, V., & Séror, J. (2007). The role of drama on cultural sensitivity, motivation and literacy in a second language context. *Journal for Learning Through the Arts, 3.*

Burton, J., Horotowitz, R., & Abeles, H. (2000). Learning in and through the arts: The question of transfer. *Studies in Art Education, 41,* 228-257.

Carter, C. S. (2005). Effects of formal dance training and education on student performance, perceived wellness, and self-concept in high school students. *Dissertation Abstracts International Section A, 65,* 2906.

Catalano, R. F., Berglund, M. L., Ryan, J. A., Lonczak, H. S., & Hawkins, J. D. (2004). Positive youth development in the United States: Research findings on evaluations of positive youth development programs. *The Annals of the American Academy of Political and Social Science, 591,* 98-124.

Catterall, J.S. (2007). Enhancing peer conflict resolution skills through drama: An experimental study. *Research in Drama Education, 12,* 163-178.

Catterall, J.S. (1998). Involvement in the arts and success in secondary school. *Americans for the Arts Monographs, 1,* 1-10.

Catterall, J.S. & Peppler, K.A. (2007). Learning in the visual arts and the worldviews of young children. *Cambridge Journal of Education, 37,* 543-560.

Center for Music Research (1990). *The Role of the fine and performing arts in high school drop-out prevention.* Tallahassee, FL: Florida Department of Education, Division of Public Schools.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* Hillsdale, NJ: Erlbaum.

Connell, J. P. (1990). Context, self, and action: A motivational analysis of self-system processes across the life-span. In D. Cicchetti (Ed.), *The self in transition: Infancy to childhood.* Chicago, IL: University of Chicago Press.

Costa-Giomi, E. (2004). Effects of three years of piano instruction on children's academic achievement, school performance and self-esteem. *Psychology of Music, 32,* 139-152.

Csikszenthmihalyi, M. & Schneider, B. (2000). *Becoming adult: How teenagers prepare for the world of work.* New York, NY: BasicBooks.

Degé, F., Wehrum, S., Stark, R. & Schwarzer, G. (2009). *Music training, cognitive abilities and self-concept of ability in children.* Proceedings of the 7th Triennial Conference of European Society for the Cognitive Sciences of Music, Jyvaskyla, Finland.

Devroop, K. (2012). The Social-emotional impact of instrumental music performance on economically disadvantaged South African students. *Music Education Research, 14,* 407-416.

DICE Consortium. (2010). *The DICE has been cast. Research findings and recommendations on educational theatre and drama.*

Dweck, C. S. (2000). *Self-theories: Their role in motivation, personality, and development.* Philadelphia, PA: Taylor & Francis.

Elpus, K. (2013). *Arts education and positive youth development: Cognitive, behavioral, and social outcomes of adolescents who study the arts.* Washington, DC: National Endowment for the Arts.

Farrington, C. A., Roderick, M., Allensworth, E., Nagaoka, J., Keyes, T. S., Johnson, D. W., & Beechum, N. O. (2012). *Teaching adolescents to become learners: The role of noncognitive factors in shaping school performance – A critical literature review.* Chicago, IL: Consortium on Chicago School Research.

Finn, J. D., & Voelkl, K. E. (1993). School characteristics related to school engagement. *Journal of Negro Education, 62,* 249-268.

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. (2004). School engagement: Potential of the concept: State of the evidence. *Review of Educational Research, 74,* 59-119.

Gourgey, A. F. (1985). The impact of an improvisational dramatics program on student attitudes and achievement. *Children's Theatre Review, 34,* 9-14.

Greene, J. P., Kisida, B., & Bowen, D. H. (2014). The educational value of field trips. *Education Next, 14.*

Heath, S. (1998). Living the arts through language and learning: A report on community-based youth organizations. *Americans for the Arts Monographs, 2,* 1-20.

Horn, J. (1992). *An exploration into the writing of original scripts by inner-city high school drama students.* Washington, DC: National Arts Education Research Center.

Kander, L. H. (2009). The impact of the Shakespearean theater production experience on the development of adolescent self-esteem, positive risk-taking and pro-social peer bonding. *Dissertation Abstracts International Section A, 70,* 37.

Kaufman, R., Rinehardt, E., Hine, H., Wilkinson, B., Tush, P., Mead, B., & Fernandez, F. (2014). The effects of a museum art program on the self-concept of children. *Art Therapy, 31,* 118-125.

Kennedy, J. R. (1998), *The Effects of musical performance, rational emotive therapy and vicarious experience on the self-efficacy and self-esteem of juvenile delinquents and disadvantaged children.* (Unpublished doctoral dissertation). University of Kansas, Lawrence, KS.

King, A. (1983). Agency, achievement, and self-concept of young adolescent art students. *Studies in Art Education, 24,* 187-194.

Lee, K. (2007). The effects of dance class on educational self-efficacy and social skills for middle school students. *Korean Association of Arts Education, 5,* 61-70

Lee, S. O. (2006). The effects of dance education on self-concept formation of high school girl students. *Korean Association of Arts Education, 4,* 55-62.

Marks, H. M. (2000). Student engagement in instructional activity: Patterns in the elementary, middle, and high school years. *American Educational Research Journal, 37,* 153-184.

Mason, M. J., & Chuang, S. (2001). Culturally-Based after-school arts programming for low-income urban children: Adaptive and preventive effects. *Journal of Primary Prevention, 22,* 45-54.

McPherson, G. E., Osborne, M. S., Barrett, M. S., Davidson, J. W., & Faulkner, R. (2015). Motivation to study music in Australian schools: The impact of music learning, gender, and socio-economic status. *Research Studies In Music Education, 37,* 141-160.

Metsäpelto, R. & Pulkkinen, L. (2012). Socioemotional behavior and school achievement in relation to extracurricular activity participation in middle childhood. *Scandinavian Journal of Educational Research, 56,* 167-182.

Minton, S. C. (2001). Assessment of high school dance students' self-esteem. *Journal of Dance Education, 1,* 63-73.

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

Peterson, C., & Seligman, M. P. (2004). Persistence [Perseverance, Industriousness]. In *Character strengths and virtues: A handbook and classification.* Washington, DC: American Psychological Association.

Rapp-Paglicci, L., Stewart, C., & Rowe, W. (2011). Can a self-regulation skills and cultural arts program promote positive outcomes in mental health symptoms and academic achievement for at-risk youth? *Journal of Social Service Research, 37,* 309-319.

Respress, T., & Lufti, C. (2006). Whole brain learning: The fine arts with students at risk. *Reclaiming Children and Youth, 15,* 24-30.

Rickard, N. S., Appelman, P., James, R., Murphy, F., Gill, A., & Bambrick, C. (2013). Orchestrating life skills: The effect of increased school-based music classes on children's social competence and self-esteem. *International Journal of Music Education, 31,* 292-309.

Roth, J., Brooks-Gunn, J., Murray, L., & Foster, W. (1998). Promoting healthy adolescents: Synthesis of youth development program evaluations. *Journal of Research on Adolescence, 8,* 423-459.

Russell-Bowie, D. (2013). Wombat stew: Enhancing self concept through an integrated arts project. *International Journal of Education & the Arts, 14,* 1-11.

Schiller, W. (2005). Children's perceptions of live arts performances: A Longitudinal study. *Early Child Development and Care, 175,* 543-552.

Scott, L. (1992). Attention and perseverance behaviors of preschool children enrolled in Suzuki violin lessons and other activities. *Journal of Research in Music Education, 40,* 225-235.

Shields, C. (2001). Music education and mentoring as intervention for at-risk urban adolescents: Their self-perceptions, opinions, and attitudes. *Journal of Research in Music Education, 49,* 273-286.

Shin, J. (2011). An Investigation of participation in weekly music workshops and its relationship to academic self-concept and self-esteem of middle school students in low-income communities. *Contributions to Music Education,* 29-42.

Smith, C., Akiva, T., Sugar, S., Lo, Y. J., Frank, K. A., Peck, S. C., & Cortina, K. S. (2012). *Continuous quality improvement in afterschool settings: Impact findings from the Youth Program Quality Intervention study.* Ypsilanti, MI: David P. Weikart Center for Youth Program Quality at the Forum for Youth Investment.

Smithrim, K. & R. Upitis (2005). Learning through the arts: Lessons of engagement. *Canadian Journal of Education, 28,* 109-127.

Tallmedge, G. K. (1977). *The joint dissemination review panel IDEABOOK.* Washington, DC: U.S. Department of Education.

Walker, E., Tabone, C. & Weltsek, G. (2011). When achievement data meet drama and arts integration. *Language Arts, 88,* 365-372.

Winner, E., & Hetland, L. (2008). Art for our sake. School arts classes matter more than ever – but not for the reasons you think. *Arts Education Policy Review, 109,* 29 – 31.

Wright, R., Lindsay, J., Ellenbogen, S., Offord, D., Duku, E., & Rowe, W. (2006). Effects of a structured arts program on the psychosocial functioning of youth from low-income communities: Findings from a Canadian longitudinal study. *The Journal of Early Adolescence, 26,* 186–205.

Zelenak, M. S. (2015). Measuring the sources of self-efficacy among secondary school music students. *Journal of Research In Music Education, 62,* 389-404.

Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory Into Practice, 41,* 64-70.

Zins, J. E., Weissberg, R. P., Wang, M. C., & Walberg, H. J. (2004). *Building academic success on social and emotional learning: What does the research say?* New York, NY: Teachers College Press.

TECHNICAL APPENDIX

SAMPLE COMPOSITION

	Overall (N = 892)		Treatment (N = 462)		Comparison (N = 430)		Difference	
	M	SD	M	SD	M	SD	t (df)	p
Age (in years)	11.93	2.85	11.54	2.42	12.35	3.21	-4.04 (793)	< .001
Gender	n	Percent	n	Percent	n	Percent	X^2 (df)	p
- Female	480	57.9	241	60.9	239	55.2	2.72 (1)	.057
- Male	349	42.1	155	39.1	194	44.8		
Ethnicity	n	Percent	n	Percent	n	Percent	X^2 (df)	p
- African American	264	32.2	139	35.6	125	29.1	19.8 (6)	.003
- Latino/Hispanic	238	29.0	103	26.4	135	31.4		
- Asian/Pacific Islander	43	5.2	29	7.4	14	3.3		
- Native American	9	1.1	4	1.0	5	1.2		
- Caucasian/White	87	10.6	38	9.7	49	11.4		
- Mixed	80	9.8	26	6.7	26	6.7		
- Other	99	12.1	51	13.1	51	13.1		

As can be seen in the table, students in the treatment group were significantly older than their comparison-group peers, and were more likely to be female at a rate approaching significance. There was also an overall effect of group on the distribution of students by ethnicity, with students in the treatment group more likely to be African American and students in the comparison group more likely to be Latino/Hispanic.

Overall 62% of students reported that they had had prior art classes in school in the same discipline (or disciplines) offered by the program in which they were participating. Students in the comparison group (65%) were more likely to have had these classes than their peers (58%) at a rate approaching significance (p = .063). Thirty-nine percent of students reported having had discipline-matched arts classes outside of school (these were defined as after-school, summer, or community-based arts classes). There were no differences by group in the proportion of students reporting that they had arts classes outside of school (p = .128).

Students' current exposure to the arts in school was measured by surveying students' primary teacher. Teachers were asked first to indicate if students in their class had or would have arts instruction in one of four disciplines (visual arts, music, dance, or theater), the number of marking periods in which that instruction would be offered (1 to 4 marking periods), and the frequency with which instruction was offered (on a four-point scale ranging from 1 = less than 1 time per week to 4 = more than 3 times per week). A score for each discipline was obtained by multiplying instruction in the discipline (0 = No, 1 = Yes) by the number of marking periods in which instruction was offered by the frequency of instruction. These discipline scores were summed to calculate an overall index of current exposure to the arts for the 2015-16 academic year. Across all students, the mean score was approximately 16, corresponding to instruction in two arts disciplines (typically visual arts and music) once a week throughout the year. There was no difference in current arts exposure by group (p = .959).

Note that this sample includes data from all students for whom parental permission and student assent were obtained, but excludes students attending programs offered by Programs D and G. Data from students served by Program D were excluded from analyses because program implementation was delayed by over three months following the collection of pre-program data. Data from students served by Program G were excluded because pre-program surveys were administered by school personnel according to a variable schedule.

Cross-Program Modeling Results, Main Effects

The table below presents the parameter estimate for group or treatment condition (1 = treatment), the model-implied (least-squared means) estimates of the post-program score for the domain indicated for the treatment and comparison conditions, and the effect size for group, which was calculated by dividing the difference in the model-implied estimates by the standard deviation (SD) of the pre-program score for a given domain. All models controlled for age, gender, race/ethnicity, and prior exposure to the arts, and all models employed a covariance structure that allowed for a random intercept for both classroom and school. Note that data collected from Programs D and G were excluded due to concerns about the schedule according to which pre- and post-program measures were administered.

Main effects for group						
Domain	Group Est.	Model-Implied Estimates			Effect Size	
	B (p)	Treatment	Comparison	Difference	SD of pre-prgm	Effect Size
Interest in Arts	.10 (.145)	3.41	3.31	.10	.872	.11
Growth Mindset	.08 (.083)	3.56	3.48	.08	.594	.13
Artistic Goal Orientation	.36 (.086)	3.26	2.90	.36	.791	.46

The Observational Measure of Quality of Instruction with Regard to SEL

Data were collected by a team of three observers who were trained to an acceptable level of inter-rater reliability α = .80. One third of observations were conducted by two observers to ensure that this level of inter-rater reliability was maintained throughout the project. Across the staff sub-scales, average inter-rater reliability was α = .86; across youth sub-scales, it was α = .83.

A pair of exploratory factor analyses were conducted: one of the staff subscales, and another of the youth subscales. Both analyses supported a one-factor solution in which all subscales loaded onto a single latent variable representing overall quality. Based on this, two scores representing overall quality of instruction with respect to SEL were calculated: one for staff, and one for youth. Internal consistency was α = .89 for the staff score and α = .84 for the youth score. These scores were correlated with one another at (r (892) = .91, p < .0001).

Modeling Results by Program

The tables below summarize the modeling results by programs. Note that only significant and sub-threshold effects are reported. The "Group Effect" refers to the p value for the variable *group*, coded 0 = comparison group, 1 = treatment group. The LSM estimates are the least-squared mean (or model-implied) estimates for each domain. All models included estimates of the variance in outcomes due to classroom. Other covariates included in each model are indicated under the effect estimates for each model.

Note that no significant or sub-threshold effects were observed for Programs I and K, and therefore no table is included for either of these programs.

Program A

Domain	Group Effect	LSM Estimates			Effect Size	
	p	Treatment	Comparison	Difference	SD, pre-program	Effect Size
Interest in Arts	.341	3.36	3.11	.25	.89	.28
Covariates: Gender, Ethnicity						
Tolerance	.462	3.74	3.60	.14	.58	.25
Covariates: Age, Gender, Ethnicity						
School Engagement	.411	3.89	3.68	.21	.79	.27
Covariates: Ethnicity						

Program B

Domain	Group Effect	LSM Estimates			Effect Size	
	p	Treatment	Comparison	Difference	SD, pre-program	Effect Size
Growth Mindset	.298	3.41	3.23	.18	.56	.32
Covariates: Age						
Academic Goal Orientation	.245	4.10	3.91	.19	.71	.27
Covariates: Age, Prior Arts in School						
Teacher Engagement	.013	3.36	2.79	.57	.92	.62
Covariates: Age						
Teacher Perseverance	.208	3.83	3.54	.29	1.05	.28
Covariates: Age, Gender, Ethnicity						

Program C

Domain	Group Effect	LSM Estimates			Effect Size	
	p	Treatment	Comparison	Difference	SD, pre-program	Effect Size
Growth Mindset	.348	3.67	3.48	.19	.61	.31
Covariates: Age						

Program E

Domain	Group Effect	LSM Estimates			Effect Size	
	p	Treatment	Comparison	Difference	p	Treatment
Teacher Engagement	.456	3.09	2.80	.29	.93	.31
Covariates: Age						
Teacher Perseverance	.514	3.76	3.46	.30	1.08	.28
Covariates: Age						

Program F

Domain	Time Effect	LSM Estimates			Effect Size	
	p	Post	Pre	Difference	SD, pre-program	Effect Size
Academic Self-Concept	.035	3.37	3.18	.19	.63	.30

Note: The evaluation of Program F employed a single-group (treatment only) design. As a result, estimates are for within-subject effects over time.

Program H

Domain	Group Effect	LSM Estimates			Effect Size	
	p	Treatment	Comparison	Difference	SD, pre-program	Effect Size
Interest in Arts	.168	3.41	2.95	.46	1.22	.38
Covariates: Ethnicity						
Tolerance	.448	3.83	3.66	.17	.46	.37
Covariates: Ethnicity						
Perseverance	.121	3.56	3.23	.33	.59	.56
Covariates: Age, Ethnicity						
Growth Mindset	.379	3.71	3.57	.14	.52	.27
Covariates: Ethnicity						
Academic Self-Efficacy	.055	3.84	3.38	.46	.65	.71
Covariates: Age, Ethnicity						

Program J

Domain	Group Effect	LSM Estimates			Effect Size	
	p	Treatment	Comparison	Difference	SD, pre-program	Effect Size
Tolerance	.588	3.70	3.54	.16	.50	.32
Covariates: Gender						
Perseverance	.643	3.49	3.33	.16	.51	.31
Covariates: None						
Growth Mindset	.287	3.37	3.22	.15	.52	.29
Covariates: None						
Academic Goal Orientation	.194	4.40	4.16	.24	.60	.40
Covariates: None						
Academic Self-Concept	.385	3.95	3.76	.19	.63	.30
Covariates: None						
Teacher Engagement	.068	3.22	2.90	.32	.82	.42
Covariates: None						
Teacher Perseverance	.185	3.64	3.27	.37	.88	.42
Covariates: Gender, Ethnicity						

Program L

Domain	Group Effect	LSM Estimates			Effect Size	
	p	Treatment	Comparison	Difference	*SD*, pre-program	Effect Size
Cultural Awareness	.240	3.98	3.75	.23	.53	.43
Covariates: Age, Prior arts out of school						
Artistic Self-Efficacy	.167	3.59	3.35	.24	.73	.33
Covariates: Age, Ethnicity						
Artistic Self-Concept	.172	3.42	3.23	.19	.67	.28
Covariates: Age						
Teacher Perseverance	.036	3.49	3.17	.32	.76	.42
Covariates: Age						

Note: The teacher measures for Program L were completed only for students in the treatment group. As a result, estimates are for within-subject effects over time.

Program M

Domain	Group Effect	LSM Estimates			Effect Size	
	p	Treatment	Comparison	Difference	*SD* of pre-prgm	Effect Size
Cultural Awareness	.480	3.95	3.77	.18	.72	.25
Covariates: Age, Ethnicity, and Prior arts in school						
Academic Goal Orientation	.224	4.04	3.62	.42	.64	.66
Covariates: Age, Ethnicity, and Prior arts in school						
Teacher Perseverance	.345	3.10	2.60	.50	.70	.71
Covariates: Age, Ethnicity, and Prior arts in school						

Cross-Program Modeling Results, Interaction Effects

Interactions with Age

Domain	Parameter	Type 3 Tests of Fixed Effects	
		F (df)	*p*
Tolerance	Age	1.27 (1, 19.2)	.274
	Gender	.95 (1, 555)	.329
	Ethnicity	.23 (6, 441)	.966
	Prior Arts in School	1.58 (1, 518)	.210
	Pre-program tolerance scores	131.7 (1, 555)	< .001
	Group	3.39 (1, 531)	.066
	Age X Group	3.48 (1, 499)	.063
Growth Mindset	Age	.09 (1, 23.5)	.770
	Gender	8.58 (1, 546)	.004
	Ethnicity	1.29 (6, 322)	.259
	Prior Arts in School	.01 (1, 452)	.911
	Group	4.54 (1, 408)	.034
	Age X Group	3.73 (1, 462)	.038
Academic Goal Orientation	Age	.48 (1, 35.4)	.492
	Gender	2.28 (1, 556)	.132
	Ethnicity	.08 (6, 491)	.999
	Prior Arts in School	1.15 (1, 504)	.285
	Group	3.17 (1, 284)	.076
	Age X Group	3.30 (1, 296)	.070

Interactions with Pre-Program Scores

Domain	Parameter	Type 3 Tests of Fixed Effects	
		F (df)	p
Perseverance	Age	.11 (1, 18.5)	.742
	Gender	.17 (1, 555)	.682
	Ethnicity	.38 (6, 368)	.892
	Prior Arts in School	2.66 (1, 474)	.104
	Pre-program perseverance scores	111.8 (1, 554)	< .001
	Group	2.92 (1, 555)	.088
	Group X Pre-program score	2.87 (1, 553)	.091
School Engagement	Age	7.88 (1, 30.1)	.009
	Gender	.14 (1, 510)	.707
	Ethnicity	1.00 (6, 472)	.428
	Prior Arts in School	4.08 (1, 467)	.044
	Pre-program engagement scores	12.5 (1, 32.1)	.001
	Group	5.71 (1, 515)	.017
	Group X Pre-program score	4.89 (1, 514)	.027
Academic Self-Efficacy	Age	.06 (1, 40)	.806
	Gender	.76 (1, 552)	.384
	Ethnicity	.52 (6, 487)	.794
	Prior Arts in School	5.11 (1, 500)	.024
	Pre-program self-efficacy scores	140.4 (1, 543)	< .0001
	Group	5.04 (1, 468)	.025
	Group X Pre-program score	4.80 (1, 492)	.029

Models Indicating a Relationship Between Arts Instruction & Socioemotional Development

The table below presents the results of a series of regressions that examined year-end scores as a function of prior exposure to the arts after controlling for students' age, gender, ethnicity, and school.

	Model Results		Arts in School			Effect Size	
	F (df)	p	B	t	p	ΔR^2	f^2
Interest	8.69 (5, 578)	< .001	.220	2.92	.004	.013	.014
Awareness	2.99 (5, 578)	.011	.169	2.63	.009	.011	.011
Self-Efficacy	5.34 (5, 577)	< .001	.161	2.83	.005	.013	.013
Self-Concept	4.25 (5, 569)	.001	.183	2.87	.004	.013	.014
Engagement	2.76 (5, 580)	.018	.144	2.19	.029	.007	.007

The following table presents the results of regressions in which year-end scores were estimated as a function of current arts instruction while controlling for these same covariates.

	Model Results		Arts in School			Effect Size	
	F (df)	p	B	t	p	ΔR^2	f^2
Self-Concept	3.63 (5, 535)	.003	.024	3.18	.002	.010	.010
Engagement	5.24 (5, 546)	< .001	.032	4.19	< .001	.030	.031

APPENDIX: PILOT STUDY REPORT

Many proponents of arts education have argued that instruction in the arts yields benefits to academic achievement, as measured by grades and standardized test scores (Winner & Hetland, 2008). However, while there is evidence that students receiving arts instruction perform better on these measures (cf. Catterall et al., 2012), much of this evidence is correlational, and therefore does not indicate that arts education *causes* higher levels of achievement. A small number of studies employing quasi-experimental designs have provided stronger evidence that arts instruction may foster cognitive development (e.g., Catterall & Waldorf, 1999). The most common explanation offered for these findings is that arts instruction fosters development in socioemotional domains necessary for both the practice of an artistic discipline and success in school (Zins, Bloodworth, Weissberg, & Walberg, 2004).

However, the specific domains and precise manner in which arts education may foster socioemotional development remains an open question. At the most general level, arts education may foster development in some of the same socioemotional domains benefitted by participation in extra-curricular activities broadly defined, including the regulation of behavior (Fredericks & Eccles, 2006) and inter-personal skills (Mahoney, Cairns, & Farmer, 2003). On the other hand, the benefits of arts education to some developmental domains may be specific to artistic modality. For example, the opportunity to participate in a program of theater education has been found to improve emotion regulation, but the same result was not realized among students receiving visual arts instruction (Goldstein, Tamir, & Winner, 2012). But it is also possible that there is a 'middle' set of developmental domains that are benefitted by instruction in arts education across disciplines, but that may be less open to change through other forms of extra-curricular activities. The ultimate purpose of the studies described in this paper is to provide a more nuanced account of whether and how arts education may foster children's socioemotional development.

Figure 1. Specificity of Benefits of Arts Education to Socioemotional Development
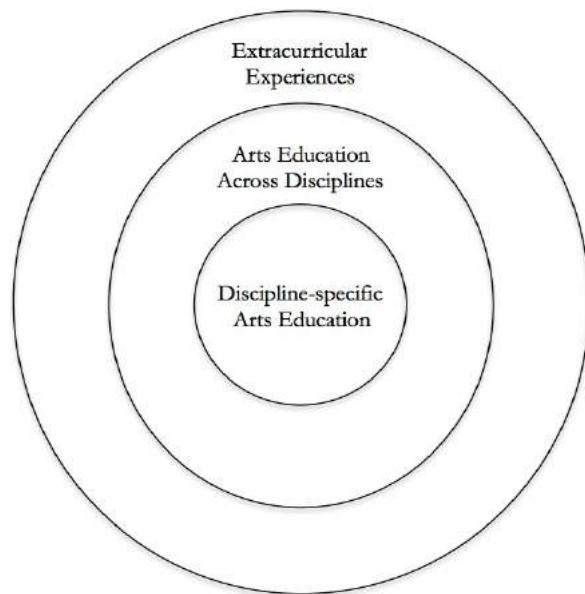
Among this middle set of developmental outcomes, those most proximal to arts education include interest in the arts, tolerance for others' perspectives, and an awareness of other cultures, and there is evidence that

even an arts education experience of very modest duration and intensity may foster development in these domains. Greene and his colleagues randomly assigned students in grades K through 12 to participate in a field trip to an art museum that included a one-hour guided tour. Students who went on the field trip evidenced higher levels of interest in the arts, tolerance, and historical empathy (empathy for people from other times; Greene, Kisida, & Bowen, 2014).

Considering more distal outcomes, the sparse research conducted to date has yielded "no more than tentative [non-correlational] evidence regarding the impact of arts education…[on] behavioral and social skills" (Winner, Goldstein, & Vincent-Lancrin, 2013, p. 19). However, this small research literature does suggest four socioemotional domains that may be influenced by arts instruction in multiple disciplines: perseverance, achievement motivation, self-efficacy, and school engagement.

- **Perseverance** refers to one's ability to exert sustained effort in the face of obstacles or in the course of difficult or lengthy tasks (Peterson & Seligman, 2004). Interest in perseverance as a character trait has been driven, in part, by its popularization as "grit," as well as findings linking perseverance to individual differences in achievement after controlling for differences in aptitude (Duckworth, Peterson, Matthews, & Kelly, 2007). Although perseverance is thought to be a relatively stable personality trait, Winner and Hetland (2008) have argued that an education in the arts may provide opportunities to engage in challenging but nevertheless rewarding tasks, and that repeated experiences of this type may foster perseverance over time (Winner & Hetland, 2008).

  However, to date only one study has yielded empirical evidence to support this argument. Scott (1992) examined rates of perseverance among preschoolers (ages 3 to 5 years) who had taken either private or group Suzuki-based violin instruction for at least five months, and compared these children to who had or had not taken creative music classes. Perseverance was assessed behaviorally during a model-building task, and was operationalized as both the amount of time students spent making the model and the degree to which their finished model resembled the example presented to them. Students enrolled in violin instruction, whether privately or in groups, scored higher in both measures of perseverance than their peers.

- **Achievement motivation** is a very broad concept and as such can be operationalized in a variety of ways. In a review of correlational links between arts education and achievement motivation, achievement motivation was defined variously as aspirations for the future, engagement in current activities, and attendance at specific events (Winner & Cooper, 2000). Thus defined, arts education has been found to correlate with higher levels of achievement motivation across multiple studies. However, a more narrow definition of achievement motivation is restricted to general, intangible achievement motives and more specific, concrete achievement goals, the latter of which are thought to have a more direct effect on achievement-related outcomes (Elliot & McGregor, 1999).

  Achievement goals may be of three types: performance-approach, performance-avoidance, and mastery goals. Of these, an orientation towards mastery goals – a desire to achieve for the sake of achievement, rather than to appear intelligent or avoid appearing foolish – have been most consistently associated with adaptive patterns of learning and achievement (Ames, 1992; Maehr, 1984; Nicholls, 1984). However, to date no study has examined change in **mastery goal orientation** as a function of arts education.

- Another aspect of achievement motivation is self-concept, which is similar to self-esteem, but is generally considered to be domain-specific: thus a bright but introverted student may have a positive academic self-concept but a negative social self-concept. Given the emphasis of this study on exploring meditational links between arts education and academic achievement, we focused our review on academic self-concept, though the literature on this topic is small. In correlational studies, both Catterall (1998) and Burton, Horowitz, and Abeles (2000) reported positive relationships between exposure to the arts and **academic self-concept**. To date, only Lee (2007) has used quasi-experimental methods to demonstrate that instruction in an art (dance) is related to higher levels of academic self-concept in a sample of school-aged students in South Korea.

- **Self-Efficacy** Like self-concept, self-efficacy is generally considered to be domain specific (Bandura, 1987). Thus a student who excels at academics but not sports may report feeling quite efficacious with regard to school but far less so with regard to basketball. Like perseverance and achievement motivation, higher levels of self-efficacy in a particular domain predict better performance in that domain, above and beyond actual capacity (Bandura, 1986).

  In a quasi-experimental study, Lee (2006) demonstrated that students attending elementary school in South Korea reported higher levels of academic self-efficacy than their peers following a program of dance instruction. Parallel results were reported by Catterall and his colleagues in two studies of students' general or global self-efficacy. In one study, Catterall and Peppler (2007) examined self-efficacy among third-grade students who participated in a program of high-quality visual arts instruction over the course of 20 to 30 weeks. Instruction was offered by two community arts organizations – Inner City Arts in Los Angeles and the Center for Contemporary Arts in St. Louis – to Latino and African American students attending public schools in those cities, and a comparison group was constructed using students attending the same grades in the same schools but who did not participate in the program. The authors found that a significantly-larger proportion of students in the treatment group (approximately half) made significant gains in self-efficacy, relative to the comparison group.

  In a second study Catterall (2007) examined the effects of an after-school drama program that was offered to students attending three middle schools in Los Angeles, the majority of whom were minorities (Latino or African American) and from low-income families. The program employed teaching artists who worked with students over the course of 24 weeks to develop a play. Students who participated in the program evidenced significantly larger gains in general self-efficacy than a comparison group of their peers recruited from the same schools.

- **School Engagement** is a multidimensional construct that includes students' participation in extra-curricular activities, feelings about school, and investment and valuation of learning (Fredericks, Blumenfeld, & Paris, 2004). Multiple researchers have proposed that school engagement predicts success in school (Connell, 1990; Finn & Voelkl, 1993), and indeed higher levels of school engagement have been found to predict lower rates of school drop out, better grades, and higher standardized test scores (Fredricks et al., 2004; Marks, 2000).

  A small number of correlational studies (Barry, Taylor, Walls, 1990; Baum & Owen, 1997; Csikszentmihalyi & Schneider, 2000) have demonstrated that students exhibit higher levels of investment and engagement in the classroom while participating in arts classes, as opposed to classes on academic subjects, but this tells us little about how the arts contribute to students' overall

engagement in school. However, two quasi-experimental studies do exist that better speak to this point. The Drama Improves Lisbon Key Competencies in Education (DICE Consortium, 2010) study was a large ($N > 5,000$), quasi-experimental study that examined the effects of drama education on a number of student outcomes, and which demonstrated that students who participated in drama education reported more positive feelings about school than their peers. In another large study, Smithrim and Upitis (2005) investigated the effects of a whole-school model of arts-integration on students' school engagement. Students in grades 1 through 6 who were attending one of 55 treatment or 35 comparison schools were asked to rate their attitudes about school at the beginning and end of the school year. Girls in the sixth grade who attended treatment schools reported significantly-higher levels of school engagement than their comparison-group peers.

- One area of socioemotional development that has not been a focus of research in arts education is **growth mindset**. Growth mindset refers to the extent to which an individual sees their intelligence as mutable, as opposed to fixed (Dweck, 1999). According to a motivational model of achievement, those who see their intelligence as open to change will be more inclined to seek out challenges and thereby maximize their potential, while those who believe their intelligence is fixed will be more likely to avoid situations in which their limitations may be exposed (Dweck & Leggett, 1999). The obvious problem with the latter approach is that avoiding challenges limits one's opportunity to improve, and thus the fixed mindset becomes a self-fulfilling prophecy. However, research has demonstrated that students' implicit theories of intelligence are open to change, and that embracing a growth mindset predicts better academic performance (Aronson, Fried, & Good, 2002). It is therefore possible that an education in the arts may alter students' theories of intelligence, as they experience firsthand a change in their own ability to appreciate or practice an artistic discipline.

In addition to the exclusion of growth mindset as a socioemotional domain that may be fostered by arts education, the literature reviewed above includes a number of limitations. First, although many of the arts education programs offered to school students are of limited duration and intensity (e.g., a field trip to a museum or attendance at a performance), few studies have examined the benefits of these programs because researchers have proceeded from the assumption that programs offering exposure to or appreciation of an art are of limited benefit (Greene, Kisida, & Bowen, 2014). While it may be unlikely that these short programs would confer benefits to certain socioemotional domains (e.g., perseverance), it is possible that they could confer benefits to other domains, such as students' interest in the arts, tolerance for others' perspectives and ideas, and an appreciation for other cultures. Second, although many researchers have argued that arts education may confer the greatest benefit to students with little previous exposure to the arts, prior exposure has typically been indexed via proxy measures such as socioeconomic status or age. Finally, few studies examine what specific pedagogical aspects of artistic instruction would benefit behavioral and social skills (see Hetland et al., 2007, for a notable exception).

Our study, to be conducted during the 2015-16 academic year, will address the following research questions:

1) Is participation in arts education programming associated with benefits to students' socioemotional development? Assuming that this is the case, does the strength of this association or the specific socioemotional domain(s) benefitted vary as a function of artistic discipline, program intensity or quality, or student age?

2) Are these benefits disproportionately realized by students who have had limited prior exposure to arts education?

3) Is there a set of identifiable features of arts programs that may differentially support socioemotional development in certain domains?

While it is not posed as a research question, we will also examine the question of transfer: the extent to which change in socioemotional development in the context of arts instruction carries over to development outside of the arts. One reason that the question of transfer is rarely addressed in studies of arts education is the complexity of the concept. As Burton and her colleagues note, it is unlikely that transfer is unidirectional (from the arts to other domains); instead, "transfer, if and where it exists, may be part of a larger constellation of impacts of arts learning on other subjects" (Burton, Horowitz, & Abeles, 2000, p. 229). Another reason is that transfer is not equally likely to be observed as a result of all programs. Indeed, it is most likely to be observed when it is a stated goal of the program. Finally, operationalizing and measuring transfer is difficult. In our study, we will assess transfer of achievement motivation (operationalized as mastery goal orientation and self-concept) and self-efficacy only among older students (those in grades 9 through 12), as these students are most likely to be able to effectively distinguish between the contexts of arts education and the more general school environment. Within the context of our design (see below), we will be able to investigate whether change in these domains within the arts predicts post-program levels in these domains outside of the arts, which would constitute evidence (albeit preliminary evidence) of transfer.

THE RESULTS OF THE PILOT STUDY

Given that we intend to employ survey methods in addressing questions one and two, and that many measures of socioemotional domains were developed with less ethnically and economically diverse samples than attend Philadelphia's schools, it was first necessary to adapt, pilot, and, if necessary, revise these measures.

As of December 2014, the Foundation had committed to supporting the work of a cohort of 15 programs in Philadelphia schools during the 2015-16 academic year. Seven programs were initially selected for inclusion the pilot study based on the following criteria:

- Whether the program would be conducting work in the schools within the timeframe of the pilot study (April through June 2015).

- Whether the program included out-of-school time (OST) activities beyond a small number of field trips or other special events.

- The ages of the students served by the program. The complete cohort of grantees included students in two broad age ranges: late-elementary school (grades 3 through 6) and high school. It was therefore important that the pilot sample include students within each of these age ranges for the purposes of measure development.

- The artistic discipline(s) covered by the program and whether the program emphasized the cultures of particular regions or traditions.

Based on these criteria, the following programs were initially chosen to be part of the pilot work:

Table 1.

Program	OST Intensive?	Age Range	Discipline(s)
Al-Bustan Seeds of Culture	Yes	Late Elementary	Music & Visual Arts*
Koresh Dance Company	No	Late Elementary	Dance
Lantern Theater Company	No	High School	Theater
Philadelphia Mural Arts	Yes	High School	Visual Arts
Philadelphia Young Playwrights	No	High School	Theater
Fleisher Art Memorial	No	Late Elementary	Visual Arts
Settlement Music School	Yes	Late Elementary	Music

Indicates the inclusion of culturally-specific programming

Following the selection process, WolfBrown engaged in a series of conversations with each program identified for inclusion in the pilot sample, with the objectives of selecting pilot study sites that: 1) program leadership were confident would be served in the 2015-16 academic year, and 2) offered the best possibility for assembling either a control or comparison group during the full study. During these conversations it became clear that running our study in parallel with the work being conducted by Dr. Eleanor Brown at Settlement Music School would be to the detriment of both projects. Therefore Settlement was removed from the pilot study sample.

Procedures by Pilot Study Site

Al-Bustan Seeds of Culture. In the 2014-15 academic year, Al-Bustan served approximately 50 students in grades 2 through 6 at Moffet Elementary School. Students were surveyed in two groups, during their music and visual arts classes. As was the case for all students in grade 6 or below, surveys were proctored by Dr. Holochwost, who read each item and all response options aloud. Eighteen students (89% female), in grades 3 through 6 ($M = 4.5$, $SD = .73$), were included as part of the pilot sample. One third of students were African American or Black (33%) or Latino/Hispanic (33%). Five students (27%) listed their ethnicity as other (four of these students indicated that they were Arabic), while one student did not provide their ethnicity.

Koresh Dance Company. Koresh served 246 students during the 2014-15 academic year at three schools: The Franklin S. Edmonds School, The General George A. McCall School, and The D. Newlin Fell School. At Edmonds, students served ranged from grade 3 to 5; at both McCall and Fell, students in fifth grade are served, along with a small number of sixth graders at Fell. Given that Edmonds offers the best possibility for assembling a comparison group during the 2015-16 academic year (see below), pilot data collection was conducted at Edmonds. Data were collected from 25 students (50% female) in grades 3 and 4 ($M = 3.5$, $SD = .51$), the majority of whom (72%) were African American. Data were collected from the students in two groups of either 12 or 13 students, with Dr. Holochwost proctoring the surveys while students' teachers were in the room.

Lantern Theater Company. The Lantern Company serves students at the Academy at Palumbo and the W.B. Saul High School of the Agricultural Sciences in grades 9 through 12. For reasons detailed below, the

Academy at Palumbo was selected for inclusion in the full study, and therefore pilot work was conducted there. Thirty students (53% female), nearly all of whom (90%) were in 9th grade and the plurality of whom were African American (50%), were surveyed during their residency, which occurs in the context of a theater elective which most students take in 9th grade. Given that these students were older, they completed the survey on their own, although Dr. Holochwost was available to answer questions students had.

Philadelphia Mural Arts Program. In the 2014-15 academic year the Mural Arts Program served high-school students at Esperanza Charter School according to a two-tier programming structure: in-school programming was offered to all students within a single grade at the school, while OST programming was offered to approximately 15 students who self-selected into the program. Due to a staff transition at Mural Arts, pilot data collection was not scheduled until late in the school year, at which point only the students in the OST portion of the program were available. This did, however, present an opportunity to administer the high school version of the survey to upper-elementary students, as Mural Arts will be serving students in 10th grade at Esperanza in the 2015-16 academic year. Twelve students (58% female) in grades 6 and 7 ($M = 6.27$, $SD = .47$) were surveyed in a single group, and were allowed to work through the survey on their own. Nearly all students were of Latino or Hispanic descent (see Table 2 below).

Philadelphia Young Playwrights. Philadelphia Young Playwrights served students at three District high schools in 2014-15: Philadelphia Military Academy, Constitution High School, and West Philadelphia High School (which may not be served in the 2015-16 academic year). At the Military Academy, the program served a single class of 10th graders, while at Constitution High School, multiple classes of students in grades 9 through 12 were served, and therefore Constitution was selected as the study site for 2015-16. Ten students (90% female) in grades 10 and 11 ($M = 10.7$, $SD = .48$) enrolled in a theater elective were surveyed, and completed the surveys on their own while Dr. Holochwost remained available to answer questions.

Fleisher Art Memorial. In the 2014-15 academic year, Fleisher worked with students at three District elementary schools: Andrew Jackson Elementary, Southwark Elementary, and Nebinger Elementary. As of this writing, Fleisher is certain to work only with students at Andrew Jackson Elementary in the 2015-16 academic year, and therefore the pilot study was conducted there. Twenty-seven students (67% female), the plurality of whom (44%) were of Latino or Hispanic descent, were surveyed while in their fourth-grade classroom (all students correctly indicated they were in fourth grade). Dr. Holochwost proctored the survey.

Table 2.

Site	N	Grade				Gender		Ethnicity					
		M	SD	Min.	Max.	M	F	AA	L/H	A/PI	White	AI/AN	Other
Al-Bustan	18	4.5	.73	3	6	2 11%	16 89%	6 33%	6 33%	0	0	0	5 28%
Koresh	25	3.5	.51	3	4	12 48%	12 48%	18 72%	0	0	0	3 12%	4 16%
Lantern	30	9.1	.58	9	12	16 53%	14 47%	15 50%	2 7%	6 20%	6 20%	0	1 3%
Mural Arts	12	6.3	.47	6	7	7 58%	5 42%	0	11 92%	0	0	0	1 8%
PYP	10	10.7	.48	10	11	9 90%	1 10%	3 30%	3 30%	0	1 10%	0	2 20%
Fleisher	27	4.0	0	4	4	18 67%	9 33%	5 19%	12 44%	4 15%	2 7%	1 4%	2 7%

Elementary	70	3.9	.60	3	6	23 33%	46 66%	29 41%	18 26%	4 6%	2 3%	4 6%	11 16%
High School	52	8.8	1.6	6	12	20 39%	32 62%	18 35%	16 31%	6 12%	7 14%	0	4 8%
OVERALL	122	6.0	2.7	3	12	43 35%	78 64%	47 39%	34 28%	10 8%	9 7%	4 3%	15 12%

Abbreviations: M (mean), SD (standard deviation), AA (African American/Black), L/H (Latino/Hispanic), A/PI (Asian/Pacific Islander), AI/AN (American Indian/Alaskan Native).
Percentages may not sum to 100% due to missing data and rounding.
The high school version of the survey was administered to a small number of 6th and 7th graders at Esperanza (see above).

Measures

Students in each program completed a packet of measures that assessed their prior exposure to the artistic discipline(s) offered by that program, their interest in that artistic discipline, their tolerance for others' perspectives, their awareness of other cultures, and each of the socioemotional domains listed above: perseverance, mastery goal orientation, academic self-concept, academic self-efficacy, school engagement, and growth mindset. In addition, older students were asked to complete items assessing their artistic self-concept and self-efficacy (in the pilot work students in grades 7 and up were asked these items, though in the full study students in grades 9 and up will be asked to complete them).

Prior exposure to the arts was assessed using items similar to those employed by Burton and her colleagues (Burton et al., 2000). Students were asked if they had taken arts classes in school, and if so, for how many years or in which grades, and whether they currently or previous participated in the arts outside of school. If they answered "yes" to the latter question, they were asked to indicate how many years that had participated in the art and what they did.

Students' **interest in the arts** was assessed using adaptations of the items used by Greene and his colleagues to gauge students' interest in the visual arts (Greene, Kisida, & Bowen, 2014). As used by Greene with students ages 6 to 18 years, these items displayed good internal reliability ($\alpha = .90$), and we did not anticipate that changing the disciplinary focus from the visual arts to include music, theater, or dance should greatly reduce reliability. Students' **tolerance for others' perspectives** and **cultural awareness** were based much more loosely on the scales developed by Greene, in part because the internal reliability of these scales was substantially lower than that reported for interest in the arts ($\alpha = .40$ and $.65$). This may have been due in part to the small number of items (3 or 4) used by Greene to assess these domains, and it may also have been attributable to younger students' difficulty in understanding abstract concepts like historical empathy. Therefore we used a larger number of items to assess tolerance and cultural awareness (items 27 – 31), and endeavored to create items that would be both germane and comprehensible to younger students. Note that as with all measures, students were asked to respond on a five-point scale ranging from "Strongly disagree" to "Strongly agree."

Perseverance was measured using the Short Grit Scale (Grit-S; Duckworth & Quinn, 2009), an 8-item reduction of the original scale, which included twelve items. The Grit-S displays high internal reliability ($\alpha = .83$), and responses to this measure have been shown to predict achievement across a number of domains. The test-retest reliability of the measure ($\alpha = .68$) across an interval of one year indicates that while

perseverance is fairly stable over time, there is sufficient inter-individual variability to allow for the possibility that perseverance may be open to change as a function of participation in arts education programs. Because the Grit-S was originally developed for use with children ages 11 to 17 years, we simplified the language of the measure prior to deploying it in our pilot work.

To measure **mastery goal orientation** we used the mastery goal subscale taken from the revised Patterns of Adaptive Learning Scales (PALS; Midgley, Maehr, Hruda, Anderman, Anderman, Freeman, et al., 2000). The PALS was developed for use with children as young as 8 years old and displays good internal reliability (α = .85), despite its brevity (5 items). Therefore only slight adjustments to the language of the original items were necessary to measure mastery goal orientation with respect to academics. However, we were also interested in measuring goal orientation in the arts among older students, and therefore we included 5 items addressing this construct as well (items 72 – 76).

One commonly-used measure of **self-concept** is the Self-Perception Profile (Harter, 2012). However, the format of the measure, which was developed to minimize positive reporting bias, is unconventional and has caused significant challenges and delays when we have used it in previous studies. Therefore we used items from the pre-adolescent version of Marsh's Self-Description Questionnaire (SDQ-I; Marsh, 1990, 1992). The SDQ-I has displayed good internal reliability (α = .88 for the general school subscale; see below), and was used successfully in Burton's study of arts education and academic self-concept, which included many students attending schools in large, urban districts. All students completed the general school or academic self-concept subscale; older students also completed a artistic self-concept subscale tailored to the discipline(s) of instruction offered by the program they attended (items 82 – 91).

As noted above, **self-efficacy** is thought to exist in both generalized and domain-specific forms. Catterall and his colleagues have investigated the relationship between arts education general self-efficacy (Catterall, 2007; Catterall & Peppler, 2007), while Lee (2006) demonstrated a relationship between arts education and academic self-efficacy. Given the focus of this study, we decided to assess academic self-efficacy among all students, but to also measure artistic self-efficacy among older students to allow a preliminary investigation into the question of transfer. Of 26 measures of self-efficacy reviewed by Rosen and his colleagues (Rosen, Glennie, Dalton, Lennon, & Bozick, 2010), only 6 were designed for use with children in elementary school, and only two of these – the PALS (Midgley et al., 2000) and a scale developed by Usher and Pajares (2008) – included subscales that were truly intended to assess self-efficacy. We chose the PALS scale, given its brevity (5 items) and acceptable internal reliability (α = .78), and included items assessing artistic self-efficacy along with those assessing academic self-efficacy (items 57 – 61) in our measure for older students.

In a recent review, Fredericks and her colleagues identified 21 measures of students' **engagement in school** (Fredricks, McColskey, Meli, Montrosse, Mordica, & Mooney, 2011). The majority of these measures asked students to report on their own levels of engagement, though many of these were focused on students' engagement in a particular class, rather than school in general. Given the goals of the study, we reviewed only those measures that asked students about their engagement in school, and focused on measures that indexed the emotional dimension of school engagement. This led us to select Voelkl's Identification with School Questionnaire (ISQ), and in particular the 6 items of the scale that exhibited the strongest relationship to the underlying construct of school belonging (Voelkl, 1996). The ISQ has exhibited acceptable internal reliability (α = [.54, .84]), and the language was sufficiently simple as to require only minor revisions.

Finally, **growth mindset** was measured using a 6-item version of the measure developed by Dweck and her colleagues (Blackwell, Trzeniewski, & Dweck, 2007). This version of the measure displayed good internal reliability (α = .77) and sufficient variability in individuals' responses, even over a short period (2 weeks test-retest reliability r = .77) to indicate that the measure should be sensitive to change in growth mindset over time. Given that the measure was designed for use with students ages 11 to 12 years, we made an attempt to simplify the language for younger students where appropriate.

Results

Over half of students (59%) indicated that they had previously taken classes in the arts discipline(s) offered by the grantees in school. On average, students reported taking these classes in 3 prior school years or grades (SD = 1.5), though this number ranged widely (from 1 to 8 years) and was correlated with students grade in school, a proxy measure of age (r (106) = .216, p = .026). A smaller proportion of the sample (41%) reported doing arts in the discipline(s) offered by the grantees outside of school. On average, students reported pursuing these activities for approximately 2 years (M = 2.2, SD = 1.6), though again estimates ranged widely (1 to 8 years), and they were correlated with grade (r (81) = .511, p < .001). Although student responses seem reasonable, it should be noted that younger students had difficulty estimating the number of years they engaged in out-of-school arts activities.

Students' **interest in the arts** was calculated as the mean of responses to items 11 – 18 for all students providing responses to at least 6 of these 8 items. Across items, data were missing for no more than 3% of the sample, and all but 8 students provided data for all 8 items (the remaining students provided data for 6 or 7 items). The mean score for the pilot sample as a whole was 3.6 out of a maximum score of 5 (SD = .81), and there was no significant difference in mean scores as a function of age group (elementary versus high school; p = .903). While there was some evidence of skew for the sample as a whole (G_1 = -.923, SE = .219), it was not severe. Internal reliability for the sample as a whole (α = .89) was similar to that reported by the measures' designers (α = .90), and did not differ as a function of age group. An exploratory factor analysis strongly suggested that all items loaded on a single factor, which accounted for 56% of the variance across items.

Tolerance for others' perspectives was indexed as the mean of responses to items 19–26 for all students providing responses to at least 6 of these 8 items, wherein items 19, 20, 23 and 26 were reverse-coded (i.e., the valence of the original response was negative, with higher scores reflecting reduced levels of tolerance). Across the sample, data were missing for no more than 4% of the sample on any item, and all but 2 students provided data for 7 (9 students) or 8 (111 students) items. For the sample as a whole, the mean score was 3.7 out of a maximum of 5 (SD = .51), with little evidence of skew (G_1 = -.588, SE = .221). Older students scored significantly higher than younger students (t (118) = -2.70, p = .008), though the difference in scores was not large ($M_{high\ school}$ = 3.9, $M_{elementary}$ = 3.6). Although internal reliability (α = .44) was similar to that reported by the measures' designers (α = .40), it was low in an absolute sense (with little difference as a function of age). Moreover, an exploratory factor analysis suggested a three-factor structure for the eight items, which stretches the definition of parsimony.

Students' **cultural awareness** was calculated as the mean of items 27 – 31 for all students answering at least 4 items. Data were missing for no more than 4% of the sample on any item, and all but four students provided data for at least 4 items. The mean score was 3.6 (SD = .76), with some evidence of skew (G_1 = -.923, SE = .219), attributable in part to a small number of outliers. Internal reliability was acceptable (α = .70), and a

single factor accounted for 47% of the observed variance in scores. There were no differences in scores or reliability by age.

Perseverance scores were calculated as the mean of items 32 through 39 for students answering at least 7 items (items 32, 34, 36, and 37 were reverse-coded prior to calculating means). All but 3 students provided data for at least 7 items, and data were missing for no more than 4% of the sample on any item. The overall mean was 3.4 (SD = .58), and there was no evidence of skew (G_1 = -.145, SE = .222). Perseverance scores were higher for students in elementary school, though this difference was not statistically significant (p = .112). Although there was no difference in internal reliability by age, overall reliability was lower than that reported by the measure's designers (α = .60). A single factor accounted for 31% of the observed variance.

Mastery Goal Orientation was assessed in the domain of academics for all students, and in the arts for students completing the high school version of the survey. **Mastery goal orientation in academics** was calculated as the mean of items 52 through 56 for all students answering at least 4 items. Data were missing for no than 4% of the sample on any item and all but three students provided data for at least three items. Although the mean score was high (M = 4.2, SD = .76) and there was clear evidence of skew (G_1 = -1.23, SE = .222), these descriptives were similar to those reported by the measure's designer. However, there was evidence of ceiling effects in students' responses (see Figure 2a). Internal reliability was identical to that reported by the designers (α = .85), and a single factor structure accounted for 63% of the variance. There were no differences in scores or reliability by age. Mastery goal orientation in the arts was calculated as the mean of items 72 – 76 for all students answering at least 4 items. Data were missing for a higher proportion students by item (12%), though all but five students provided data for at least 4 items. Although scores were not as high and skew was not as pronounced as in the case of academic mastery goal orientation (M = 3.7, SD = 1.2, G_1 = -.751, SE = .347), ceiling effects were still observed (see Figure 1b). The internal reliability of the measure was high (α = .96) and a single factor accounted for 88% of the observed variance. The correlation between mastery goal orientation in academics and the arts was moderate and statistically significant (r (52) = .318, p = .029).

Figure 2a. Mastery Goal Orientation in Academics   Figure 2b. Mastery Goal Orientation in the Arts
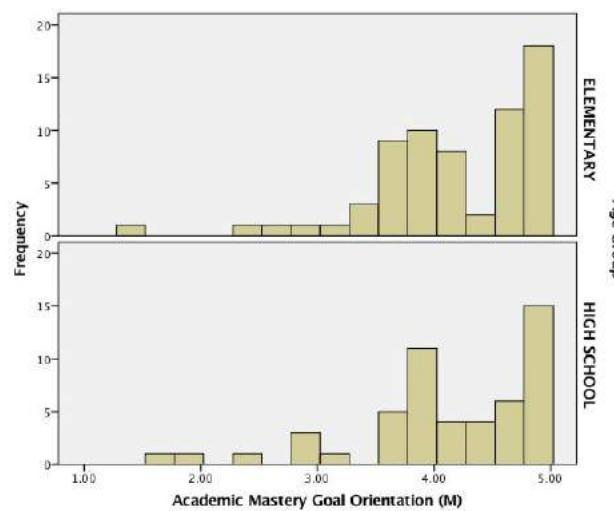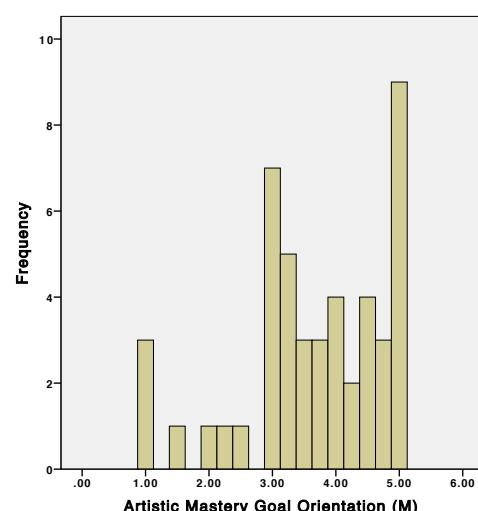

Like goal orientation, **self-concept** was measured in the domains of academics and the arts. **Academic self-concept** was measured as the mean of items 62 – 71 (with items 65 and 68 reverse coded) for students

answering at least 8 items. Data was missing for no more than 7% on any item, and all but 4 students provided data for at least 8 items. The mean score was 3.6 (SD = .72), with no evidence of skew (G_1 = -.355, SE = .223). Internal reliability, while slightly lower than that reported by the measure's developers, was good (α = .80), and a single factor accounted for 38% of the variance in scores. There were no significant differences in scores as a function of age. The measure of **Artistic self-concept** displayed similar descriptives (M = 3.7, SD = 1.2, G_1 = -.626, SE = .350), internal reliability (α = .91), and factor structure, with a single factor accounting for 56% of the variance in scores. Scores for this domain were calculated as the mean of items 82 – 91 (85 and 88 reverse coded) for students answering at least 8 items. Data were missing for no more than 13% of students on any item, and all but 6 students provided data for at least 8 items. Academic and artistic self-concept scores were significantly correlated (r (50) = .402, p = .006).

Self-efficacy was the third and final construct assessed in the domains of both academics and the arts. **Academic self-efficacy** was assessed as the mean of items 57 – 61 for students answering at least 4 items. Data were missing for no more than 5% of students on any item, and all but 3 students provided data for at least 4 items. The mean score, while high (M = 4.0, SD = .70), was similar to that reported by the measure's designers, as was the skew statistic (G_1 = -1.04, SE = .222). Internal reliability was slightly higher than that reported by the measure's designers (α = .81), and a single factor accounted for 58% of the variance in scores. There were no significant differences in scores as a function of age. While both descriptives (M = 3.9, SD = .92, G_1 = -.844, SE = .340) and internal reliability (α = .90) were similar for the measure of **artistic self-efficacy**, for this measure there was clear evidence of ceiling effects (see Figure 3). Artistic self-efficacy was calculated as the mean of items 77 – 81 for all students answering at least 4 items. Data were missing for no more than 8% of students on any item, and all but 3 students answered at least 4 items. There was a moderate correlation between academic and artistic self-efficacy scores (r (51) = .452, p = .001).

Figure 3. Artistic Self-Efficacy
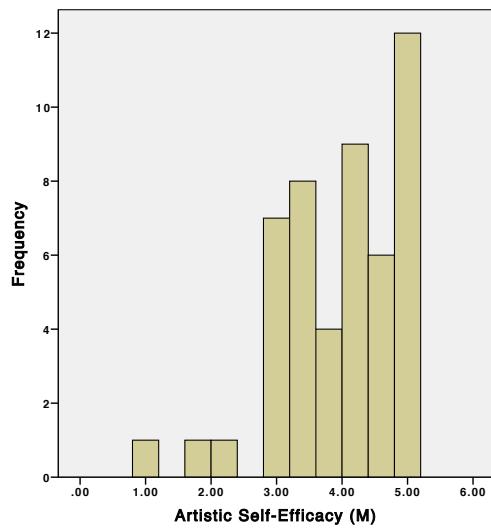
Students **school engagement** was measured as the mean of items 46 – 51 (with items 47, 48, and 49 reverse coded) for all students answering at least 5 items. Data were missing for less than 4% of students on any item, and all but 4 students provided data for at least 5 items. The overall score on the measure was 3.6 (SD = .72) and there was no evidence of skew (G_1 = -.300, SE = .223). Scores did not differ significantly by age. Internal reliability, while modest (α = .62), was within the range of what has been reported by others using the measure. A single factor accounted for 35% of the variance in scores.

Finally, **growth mindset** was calculated as the mean of items 40 – 45 (items 41, 42, and 44 reverse coded) for students answering at least 5 items. Data were missing for no more than 5% of students on any item, and all but 4 students provided data for at least 5 items. The overall score was 3.6 (*SD* = .83) with no evidence of skew (G_1 = .048, *SE* = .223), though scores were higher among high score students to a degree approaching significance (*p* = .064). There was also evidence of ceiling effects, particularly among younger students (see Figure 4). Internal reliability was identical to that reported by the measure's designer (α = .77) and a single factor accounted for 47% of the variance in scores.

Figure 4. Growth Mindset
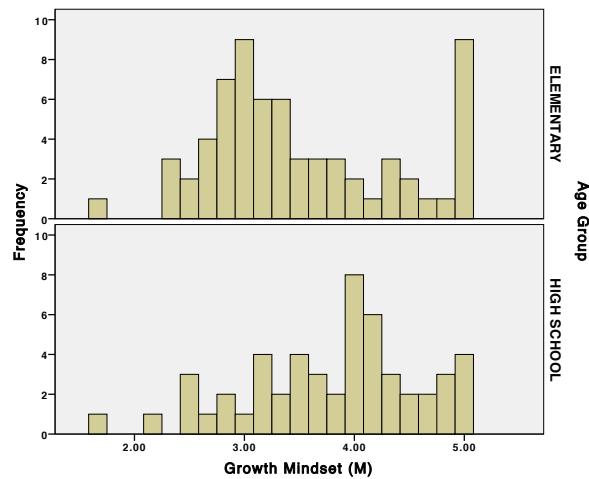
Summary of Results and Implications for Full Study Design

Table 3 presents a summary of the psychometric properties for the measures used in the pilot study.

Table 3.

Domain	*M*	*SD*	Range		Skew		Reliability (α)
			Minimum	Maximum	G_1	*SE*	
Interest in the Arts	3.6	.81	1.0	4.9	-.923	.219	.89
Tolerance	3.7	.51	2.0	4.8	-.588	.221	.44
Cultural Awareness	3.6	.76	1.0	5.0	-.742	.223	.70
Perseverance	3.4	.58	1.8	5.0	-.145	.222	.60
Academic Goal Orientation	4.2	.76	1.4	5.0	-1.23	.222	.85
Artistic Goal Orientation	3.7	1.2	1.0	5.0	-.751	.347	.96
Academic Self-Concept	3.6	.72	1.8	5.0	-.355	.223	.80
Artistic Self-Concept	3.7	.92	1.2	5.0	-.626	.350	.91
Academic Self-Efficacy	4.0	.70	1.2	5.0	-1.04	.222	.81
Artistic Self-Efficacy	3.9	.92	1.0	5.0	-.844	.340	.90
School Engagement	3.6	.72	2.0	5.0	-.300	.223	.62
Growth Mindset	3.6	.83	1.7	5.0	-.048	.223	.77

Based on this summary, the more detailed pilot results, and our experience in administering the pilot measures, we will make the following revisions:

- We will remove the question asking younger students to estimate the number of years they have participated in the arts outside of school, and will make certain to assist students in counting the number of previous grades in which they may have received in-school instruction.

- We will add a small number of items that deliberately set a high threshold for a positive response (or a negative response, in the case of reverse-coded items) to our measures of goal orientation (academic and artistic), artistic self-efficacy, and growth mindset, in an effort to reduce the ceiling effects observed for the responses to these measures.

Note that we will not alter the measures of tolerance or perseverance based on the modest estimates of internal reliability.

THE DESIGN OF THE FULL STUDY

Although the paper-and-pencil measures performed well, these measures have limitations, both minor and major, that necessitate the inclusion of additional measures in the full study. Self-report is the only valid way to collect data on for many of the domains of socioemotional development included in our study. Mastery goal orientation, self-concept, self-efficacy, and growth mindset are inherently a function of how students view themselves, and therefore can only be assessed by self-report. However, these measures are subject to acquiescence bias, which refers to the tendency of individuals' self-reported characteristics to drift upwards upon repeated administration. Therefore, within the limits imposed by constraints on teachers' time, students' in-school teachers will be asked to complete the informant version of the short grit scale (Duckworth & Quinn, 2009) and a three-item measure of student engagement in school from the Research Assessment Package for Schools (RAPS; Institute for Research and Reform in Education, 1998) for each child participating in the study.

But even with the addition of these teacher measures, observational measures are required in order to answer the research questions posed for the study. While data collected by the Foundation allows us to speak to program intensity or dosage as defined in our first research question, assessing the quality of programs as implemented will require an observational measure of program quality broadly defined. Answering our third research question – whether there are identifiable features of arts programs that may differentially support socioemotional development across domains – requires a more finely-grained observational measure of the curricular or pedagogical content of these programs.

To this end, a complementary part of the pilot study work consisted of field-testing and adapting the Youth Programming Quality Assessment (YPQA; Smith et al., 2011). The YPQA is an observational measure that assesses the quality of program implementation across six dimensions: 1) safe and supportive environment; 2) interaction; 3) engagement; 4) youth-centered policies and practices; 5) high expectations for youth and staff; and 6) access. As part of a larger study of improving the quality of out-of-school programming for youth (The Youth Program Quality Intervention Study; Smith et al., 2012), pairs of trained observers used the YPQA to assess 32 different afterschool programs and obtained an inter-rater reliability of $\alpha = .61$.

To this we added a seventh dimension regarding creative habits, which was designed to assess the degree to which imagining different possibilities, reflection and critical thinking, courage and risk-taking, persistence and discipline in the context of arts education were featured in program practice. The revised YPQA, including the creative habits scale, was piloted with programs of arts and arts-integrated instruction administered by Arts Corps in Seattle, Washington. Piloting revealed inter-rater reliability for pairs of observers that was acceptable ($\alpha \approx .75$) but that will need to be improved through additional training in the fall.

The revised YPQA offers a number of advantages for assessing the quality of programs. First, the base measure on which it is based is widely used in a variety of youth programs. Second, in its revised form it is broadly-consistent with other rubrics for quality in arts education programming (e.g., Seidel et al., 2009). Third, and perhaps most importantly, by offering a multidimensional assessment of program quality, it allows for the possibility that programs may excel in a variety of ways. Not only does this facilitate the acknowledgement of a program's strengths, but it also allows for the possibility that different dimensions of quality may more strongly predict growth in specific socioemotional domains. Together with other data on program characteristics and student factors, the data yielded by the revised YPQA affords us the opportunity

to investigate how these characteristics and factors may combine to predict students' socioemotional development, as depicted in Figure 5.

Figure 5. A Conceptual Model of Socioemotional Development
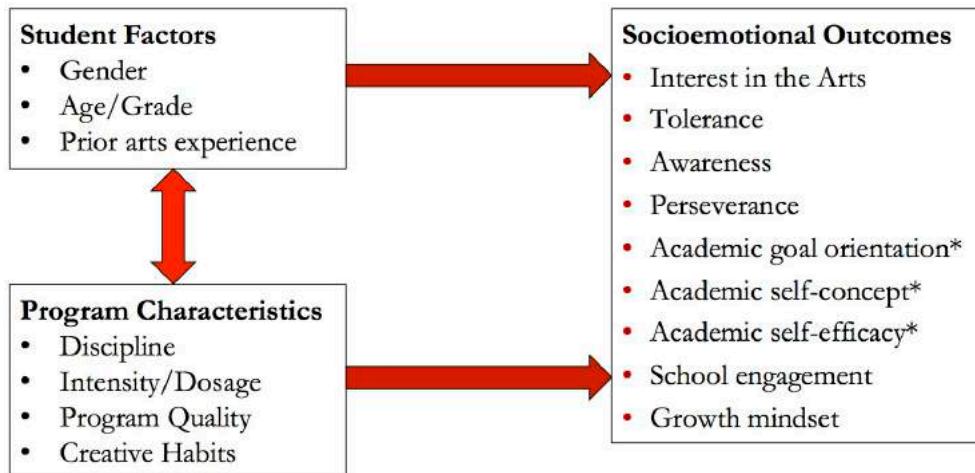
Overview of Procedures

The observational measures will be collected during regular program practice, during two observations of each program (with the exception of assembly-based programming; see below). These observations will be scheduled to occur no earlier than mid-way through the program, and will be conducted to capture as broad and representative an account of program practice as possible. All observations will be conducted by trained observers using a scoring rubric and accompanying codebook.

For the paper-and-pencil measures completed by students and teachers decisions about who completes the measures and when will dictate which design is employed, of which there are four options:

- Two programs can be studied using a true **experimental design**. At New Foundations Charter School school leadership will assign approximately 50 students at random to receive the program in the 2015-16 academic year, while holding another 50 students who elected to receive the program out to comprise a control group. Because Opera Philadelphia's Opera in the Classroom Program consists of a single 45-minute workshop that all students receive, an abbreviated set of measures (assessing prior exposure, interest, tolerance, and cultural awareness) will be administered to students just before (for the randomly-selected control group) or just after the workshop.

- Four programs included in the full study are amenable only to a **quasi-experimental design** featuring a matched comparison group. Al-Bustan Seeds of Culture, Koresh Dance Company, Philadelphia Mural Arts, and Fleisher Art Memorial all offer programming to students over the majority of the entire school year (e.g., from October to May), but serve only a subset of students at a given school in selected grades. Therefore students whose parents provide permission for them to participate in the study and who are enrolled in the program will constitute the treatment group, while their grade-matched peers who are not enrolled in the program will constitute the comparison group (where possible, members of groups will be matched by teacher as well). Both

groups will complete the full set of age-appropriate measures as close to the beginning and end of the program as possible.

▪ The remaining programs run for only a portion of the school year, and of these, four programs – Astral Artists, Musicopia, Spiral Q Puppet Theater, and the Walnut Street Theater – follow a whole-grade implementation model, which precludes the possibility of assembling a comparison group. However, because these programs do not begin their programming until January, it would be possible to study them using an **experimental design with delayed implementation**. Under this design, all permitted students will be assigned at random to either a control or treatment group, after stratifying by classroom. Students in the control group will complete their measures prior to their participation in the program, with pre- and post-administration of the measures separated by an interval equal to the length of the program. Students in the treatment group will complete their measures concurrent with the beginning and end of the program. While this design guards against certain threats to internal validity (e.g., selection bias), it does conflate program and developmental effects: it is possible, for example, that rates of change in various socioemotional domains would vary across the course of the school year, independent of program effects.

An alternative would be to employ an **interrupted time-series design**. In this design all permitted students would complete the measures three times: the second and third administrations of the measures would be concurrent with the beginning and end of the program, but the first administration would precede the beginning of the program by an interval equal to the program length. Using this design we could examine whether change in socioemotional domains from the second to the third administrations of the measures exceeded change observed from the first to the second administrations. However, we could not rule out the possibility that an increased rate of change would have been observed absent the program, as no comparison group would be available. One stronger elaboration on this design would extend the period of time covered to include the following school year, and collect six data points for each child: data points A_1 and A_2 would be collected in the school year that students' received programming, but prior to program implementation, separated by an interval of time equal to the length of the program; data points B_1 and B_2 would be collected concurrent with the beginning and end of the program, respectively; data points A_3 and A_4 would be collected in the school year following students' enrollment in the program, on a schedule analogous to that used to collect points A_1 and A_2. The logic of this design is similar to that employed in many single-case designs that feature the absence, administration, and subsequent absence of an intervention or treatment. The essential hypothesis is that growth in relevant socioemotional domains during the time span coinciding with the program (B_1 to B_2) would outpace growth observed both prior to (A_1 and A_2) and following (A_3 and A_4) the program.

The last group of programs – Lantern Theater Company, Philly Young Playwrights, Taller, and the Wilma Theater – run for only a portion of the school year, they do not serve all students in a grade. This raises the possibility of employing any of the designs described above (with the exception a true experimental design) including a **hybrid design**, in which students a portion of the students who will be in the program constitute a control group within an experimental design with delayed implementation, while students who will not receive the program constitute a comparison group in a quasi-experimental design. Note, however, that for each these four programs program and teacher effects would be conflated since students are assigned to the program based on classroom.

In choosing a design, it is also important to consider the issue of acquiescence bias. The shorter the interval of separation between repeated administrations of a measure, and the more times a measure is

administered, the more acquiescence bias is generally observed. A control or comparison group can aid in guarding against acquiescence bias: assuming that acquiescence bias effects students in the treatment and control or comparison groups equally, we will be able to partial out the change in ratings of socioemotional development that is genuinely attributable to the influence of the program.

The section below outlines how each program will be implemented at the study sites, and presents the design that will be employed in the full study. Two design alternatives are presented for the final two groups of programs.

Procedures by Program

Experimental Design

Rock Reach School. In the 2014-15 academic, the Rock Reach School served students at 10 schools: 5 District neighborhood schools, 4 charter schools, and 1 parochial school. At three of these schools (John Wister, Anna B. Day, and Mastery Charter School), Rock Reach serves students outside of the age range of the current study, while at three other schools, students enter the program through audition (Khepera), satisfying certain criteria (U.S. Dream Academy), or mandatory assignment (Maternity BVM), making these less desirable sites for study. At Nebinger Elementary and Global Leadership Academy students (or their families) elect to participate in the program, while at W.C. Bryant Promise Academy the program is offered as part of the curriculum for students in grades 3 through 5, raising the possibility of assembling a comparison group of non-selecting students (at Nebinger and Global) or non-assigned students (at Promise Academy).

However, at New Foundations Charter School students in grades 3 to 5 choose their top four electives, including Rock Reach's program. School leadership is willing to assign approximately 50 students at random to receive the program in the 2015-16 academic year, allowing for the assembly of a true control group of approximately 50 students. Program leadership have agreed to maintain these waiting lists for the 2015-16 academic year, and to prioritize enrollment in the program for the 2016-17 school year for students participating in the study.

Opera Philadelphia. As of this writing, a number of schools have expressed their interest in having Opera Philadelphia offer their Opera in the Classroom Program to their students during the 2015-16 academic year. Opera Philadelphia selected five schools with which they are most confident they will work during the 2015-16 season, based on their experience during the 2014-15 academic year. Three of these schools serve students in the age range of our study: H.A. Brown Elementary (150 students in grades 3 through 8), Rawnhurst Elementary (150 students in grades 4 and 5) and Webster Elementary (125 students in grade 4).

The brevity of the Opera in the Classroom Program would allow for the assembly of a control group. Once it is known which classrooms at a given school would receive the program, children in these classrooms would be stratified by classroom and then randomly-assigned to either the treatment or control group. Students in the control group would complete measures prior to receiving the program, while students in the treatment group complete measures after receiving the program. Given that the program consists of a single 45-minute workshop, a reduced set of measures, featuring our measures of prior exposure to the arts, interest, tolerance, and cultural awareness would be administered once to students in both groups.

Quasi-Experimental Design with Matched Comparison Group

Al-Bustan Seeds of Culture. In the 2015-16 academic year Al-Bustan will continue to work with students in grades 2 through 6 at Moffet Elementary, but will seek to increase enrollment to approximately 75 students.

Students' families select to have their children participate in the program, and the program runs for nearly the entirety of the school year. There are approximately 225 students in the age range of our study (grades 3 through 6) at Moffet Elementary who are not enrolled in the program, and these students will be invited to participate in the study as members of the comparison group. Data collection from both groups will coincide with the beginning and end of Al-Bustan's 2015-16 program year, which runs from September through May. Note that Al-Bustan's program consists of three modes of participation: an introductory music program, an advanced music program (a percussion ensemble), and a visual arts program. Students' enrollment in these programs will be tracked to allow for the investigation of differential program effects by mode of participation.

Koresh Dance Company. Edmonds Elementary School offers four advantages over McCall and Fell Elementary schools as a site for the full study: 1) Edmonds is the school with which Koresh has worked for the longest period of time, and therefore represents program implementation at its most developed; 2) Koresh serves the largest number of children at Edmonds; 3) Only at Edmonds are there classes in each grade that are not served by the program; and 4) Only at Edmonds can we investigate whether the effects of the program vary as a function of the number of years a student has participated. All students in the two 3rd grade classrooms, one 4th grade classroom, and one 5th grade classroom served by the program will be invited to participate in the study as members of the treatment group, while students in the single 3rd grade classroom and two 4th and 5th grade classrooms not served by the program will be invited to participate as members of the comparison group. It is also of note that at Edmonds neither teachers nor students self-select into the program; rather, the Principal assigns classrooms to receive the program based on whether those classrooms are free when Koresh is at the school. Therefore, while students who do not receive the program are not a randomly-assigned control group, the issues of selection bias are reduced. Data collection will occur in October 2015 and May 2016.

Philadelphia Mural Arts Program. Data collection for the full study will include students in the in-school portion of Mural Arts' programming, as too few students participate in the OST to allow recruitment of a sufficient sample size. Data collection will include students who are enrolled in 10th grade at Esperanza Charter School. Approximately 75 of these students (those enrolled in three academic subjects, as of yet unknown) will receive the program; the remainder of students in 10th grade (totaling approximately 125 students) will not. Students who receive the program and consent to participate will comprise the treatment group, while the remainder of the 10th graders will comprise the comparison group. Data will be collected from students in both groups concurrent with the beginning and end of Mural Arts' program year, which runs from October to June.

Fleisher Art Memorial. As of this writing, Fleisher was certain to work with third- and fourth-grade students at Andrew Jackson and Southwark Elementary schools during the 2015-16 academic year. At each school, there are two classrooms in each grade, and all classrooms include approximately 25 students. Students' teachers elect to have their class participate in the program; students enrolled in these classrooms within each grade at each school will comprise the treatment group, while those in the classrooms not participating in the program will comprise the comparison group. Data collection will occur concurrent with the beginning and end of the program, though as of this writing beginning and end dates have not been determined.

Experimental Design with Delayed Implementation or Interrupted Time Series Design
Astral Artists. In the 2014-15 academic year, Astral Artists began working with students in all third grade classrooms at three schools: 30 students in a single classroom at Chester A. Arthur, 70 students in two

classrooms at Andrew Jackson, and 70 students in two classrooms at Moffet Elementary. In the 2015-16 academic year, Astral will continue to work with fourth-grade students who they served the previous year as third graders, as well as any students who are new to fourth grade. Each student will participate in 6 classes beginning in January 2016 and ending in April. Under an experimental design with delayed implementation all fourth-grade students at Jackson and Moffet (the two schools where the largest number of students are served) would comprise both the treatment and control groups – after stratifying by fourth-grade classroom teacher, students will be assigned at random to one of these groups. Students assigned to the control group will complete their measures in the fall of 2015, with pre- and post-measures occurring approximately 10 weeks apart. Students in the treatment group will complete their measures prior to and after participating in the program. Under an interrupted time-series design, data collection would include all students who would receive the program, and all students would be surveyed three times: ten weeks prior to program implementation, at program implementation, and at the program's conclusion.

Musicopia. Musicopia's Dancing Classrooms program occurred at five schools during the 2014-15 academic year: the General Meade School, the Henry Houston School, and the Eugenio De Maria de Hostos Charter School, Philadelphia Montessori Charter School, and Visitation BVM. With one exception (Meade), the program served students in 5th grade, and in every case all classrooms within a grade were served. As of this writing, Musicopia anticipates serving 5th grade students at de Hostos Chater and Visitation BVM in the 2015-16 academic year; data collection will occur at de Hostos (Visitation BVM is not covered under the District's Research Review Committee), where approximately 80 students in four classrooms will be served. Students will receive two lessons per week for 10 weeks, beginning in January 2016.

Under an experimental design with delayed implementation, students would be 5th grade students will be assigned at random to the treatment or control condition after stratifying by classroom. All students would eventually in the program; however, students in the control condition will complete measures approximately 10 weeks apart in the fall of 2015, prior to program participation. Students in the treatment condition will complete measures concurrent with their participation in the program. Alternatively, under an interrupted time-series design all students would be surveyed according to the schedule outlined for Astral Artists.

Spiral Q Puppet Theater. In the 2015-16 academic year, Spiral Q will serve students at five schools through their residency and Ignition programs: Belmont Charter School, Samuel Rhoads Elementary, Rudolph Blankenburg Elementary, Middle Years Alternative School, and Alain Locke Elementary. At each school they served approximately 60 children through their residency (all sixth graders with the exception of Rhoads Elementary, where they served fifth graders) and between 250 and 600 students through their Ignition program. All students in a grade receive the residency, so assembling a within-grade comparison group of students is not possible, and all students at the school receive the ignition program. However, at Middle Years Alternative School program leadership believe a delayed implementation of the program should be possible, which would allow the use of an experimental design with delayed implementation or an interrupted time-series design.

Walnut Street Theater. In the 2015-16 academic year, the Walnut Street Theater will be serving students at the Pan American Academy through its Adopt a School program. Like Mural Arts and Spiral Q, the Adopt a School initiative has two levels of intensity: an in-school component featuring a theater integrated lesson plan that runs for 6 weeks and is offered by two teaching artists to 12 sections of students, in which all students participate, and an after-school program into which two groups of 15 students self-select. Given that the number of students participating in the after-school program is small, the study will focus on the

in-school programming. The implementation of this programming is staggered, with students in grades 3 through 8 receiving the program in one of five sessions: 1) mid-September through early October, 2) mid-October through early December, 3) early December through mid-January, 4) mid-January through late March, and 5) early April through mid-May. In 2015-16, students in grades 3 through 5 will receive the program in sessions 2 and 3, while students in grades 6 through 8 will receive the program in sessions 4 and 5.

Under an experimental design with delayed implementation, all students who consent to participate in the study would be assigned at random to either the treatment or the control group. Students in grades 3 through 5 who were assigned to the control group would complete their measures in session 1; students in these grades who are in the treatment group would complete their measures concurrent with their participation in the program; students in grades 6 through 8 who are assigned to the control group would complete their measures in session 3, while those in the treatment group would complete their measures concurrent with the program. Alternatively, data collection could be re-arranged consistent with an interrupted time-series design, with students completing measures three times, 6 weeks apart, with the second and third collections coinciding with the beginning and end of the session in which they received the program.

Experimental Design with Delayed Implementation or Hybrid Design
Lantern Theater Company. One of the two schools served by Lantern, Saul High School, is too small to allow for the recruitment of a comparison group matched by grade. At the Academy at Palumbo, however, Lantern serves three classrooms of 9th graders taught by a single teacher and four classrooms of 10th graders taught by another teacher, but does not serve 6 additional classrooms in 9th grade and four additional classrooms in 10th grade. Moreover, in the 2015-16 Lantern's programming will occur late in the school year (between March and May, 2016), making it possible to recruit both a treatment and a control group from the students in the classrooms Lantern will serve. Under an experimental design with delayed implementation, consenting students would be assigned to these groups at random; those in the control group would be surveyed early in 2016, with pre- and post- data separated by an interval approximately equal in duration to the length of the program (approximately 8 weeks). These students would be free to participate in the program when it occurred, but they would not be surveyed again. Instead, only those students in the treatment group would be surveyed concurrent with the beginning and ending of Lantern's programming. Under a hybrid design, the sample could be supplemented with a comparison group of selected from students in the classrooms Lantern does not serve, who would be surveyed concurrent with the beginning and end of the program.

Philadelphia Young Playwrights. Data collection for the full study will be conducted at Constitution High School, as this site offers robust numbers of students in both the treatment and comparison groups (as opposed to Philadelphia Military Academy) and Young Playwrights is certain they will work in this school during the 2015-16 academic year (unlike West Philadelphia High School). The treatment group will be defined as students enrolled in elective classes that receive the residency (~100 students); students enrolled in the same grades in classes that do not receive the residency will comprise the comparison group (~75 students).

Taller Puertoriqueno. In the 2014-15 academic Taller worked with students at Potter Thomas Elementary School, William Cramp Elementary, Moffet Elementary, The Luis Muñoz Marín School, Isaac Sheppard Elementary, and Julia de Burgos Elementary School. As of this writing, the program is most confident that it will continue to work with students at Cramp, Moffet, and Sheppard Elementary schools, but only at Cramp

will it serve multiple classrooms of students in grades 3, 4, and 5. The leadership at Cramp is working with Taller to schedule the implementation of programming by classrooms throughout the school year, and as of this writing program implementation is likely to be staggered by classroom. This would allow for the recruitment of a control group from among students who will ultimately receive the program but have not yet done so, but also allowing for the use of a hybrid design.

Wilma Theater. In the 2014-15 academic year, the Wilma Theater served students at 30 students at Bartram High School, 120 students at Carver High School of Engineering and Science, and 15 students at El Centro de Estudiantes through its Wilmagination program, which featured 14-session residencies that occurred over the course of approximately 8 weeks. In the 2015-16 academic year Wilma anticipates working with Constitution High School, John Bartram High School, and Carver High School, where they will serve specific classrooms within grades at each school. At Carver they will work with two English classrooms of 12th graders taught by the same teacher from mid-March to mid-May. Each of these classrooms comprises approximately 30 students, and there are two similarly-sized classrooms of students (taught by another teacher) who will not receive the program. Therefore both an experimental design with delayed implementation or a hybrid design are feasible.

Note that it is possible that students at a small number of schools will have participated in multiple programs. At Andrew Jackson Elementary, students in 4th grade may receive programming from both Fleisher and Astral Artists. At Moffet Elementary students in grades 3, 4 or 5 may participate in some combination of Al-Bustan's, Astral Artists', and Taller's programming, while at Constitution High School, students may receive programming from both Philadelphia Young Playwrights and the Wilma Theater. Students who are receiving programming from multiple grantees will participate in data collection with the grantee for which data collection begins first, and will not participate in data collection with programs for which data collection begins thereafter. However, the receipt of multiple programs by a single student will be noted and included as a variable in our subsequent analyses.

Data Analyses

Data analyses will proceed in two steps, and all analyses will be conducted within a hierarchical modeling framework, given that the data for most programs are likely to be nested by in-school teacher, and school. In the first step of our analyses, we will address whether participation in arts education programming is associated with benefits to students' socioemotional development. With the exception of Opera Philadelphia, the analyses for programs studied under experimental (both with or without delayed implementation) or quasi-experimental designs will examine post-intervention scores after controlling for pre-intervention scores. The resulting parameter estimate for group status (treatment versus control or comparison) will indicate the magnitude of the differences in post-intervention scores after holding pre-intervention scores constant. Interrupted time-series designs will add a third data point, requiring that models be re-parameterized so that time is nested within child and time becomes the focal predictor, with the question of interest being how the slope representing change in socioemotional development differs between data points 1 and 2, as compared to data points 2 and 3.

In the second step of our analyses we would add additional independent variables to our models. These would include program factors – the artistic discipline of the program, its intensity, and its quality – as well as student factors: gender, age, and prior exposure to the arts. This would allow us to address whether the strength of the association between participation in arts education or the specific socioemotional domain(s) benefitted varied as a function of artistic discipline, program intensity or quality, or student age, as well as

whether any benefits were disproportionately realized by students who had limited prior exposure to arts education.

Conclusion

This report is a technical document, in that it focuses heavily on the topics of measure selection, development, and eventual deployment. We deemed this focus appropriate for this stage in the project, when deficiencies in measurement or design can still be corrected. Given the thoughtful feedback we have already received from our colleagues at the Foundation, we decided to provide a degree of technical detail that would allow the reader to understand the multitude of decisions made to date and our rationale for making those decisions. Unfortunately, this level of detail also limits the broader appeal of the report, which is not intended as a public document in its current state.

There is, however, an extension of this paper that may have wider appeal. Conducting the pilot study and planning the full study has reinforced the challenges of conducting research in a setting like the Philadelphia public schools, which include disruptions of communications and scheduling, high rates of absence among students and high rates of turnover among teachers and staff. These challenges should not become excuses for conducting poor work, and therefore we have outlined a rigorous research design for the full study, but with the understanding that this design is to some extent aspirational: the realities on the ground at each site ultimately determine the feasibility of our design as implemented.

Although it is rarely acknowledged, this tension between rigor and reality exerts great influence on what research is conducted and with whom. In the interest of being able to employ rigorous designs researchers often gravitate towards more stable schools and districts that disproportionately serve more affluent students. As a result, their findings are less generalizable to students from backgrounds similar to the students attending Philadelphia's public schools. An extension of this white paper would begin by acknowledging this issue and its consequences, which include the fact that the evidence base in support of arts education is weakest for precisely those students who are least likely to receive it. The paper could then outline designs that afford the optimum balance of rigor and feasibility in challenging contexts, drawing upon lessons learned in conducting both the pilot and full studies. This would constitute another contribution to the field on the part of this project and the Foundation.