



---

# **Evaluation of Professional Development in the Use of Arts-Integrated Activities With Mathematics Content: Findings From the Evaluation of the Wolf Trap Arts in Education Model Development and Dissemination Grant**

**Final Grant Report**

**Meredith Ludwig, Ed.D.**

**Mengli Song, Ph.D.**

Submitted to Wolf Trap Foundation for the Performing Arts

JANUARY 2015  
(UPDATED FEBRUARY 2016)



**Evaluation of Professional Development in  
the Use of Arts-Integrated Activities With  
Mathematics Content:  
Findings From the Evaluation of the Wolf Trap  
Arts in Education Model Development and  
Dissemination Grant**

**Final Grant Report**

**January 2015**

**(Updated February 2016)**

**Meredith Ludwig, Ed.D.**

**Mengli Song, Ph.D.**



1000 Thomas Jefferson Street NW  
Washington, DC 20007-3835  
202.403.5000 | TTY 877.334.3499

**[www.air.org](http://www.air.org)**

Copyright © 2016 American Institutes for Research. All rights reserved.





The contents of this publication were developed under a grant from the U.S. Department of Education. The views expressed herein, however, do not necessarily represent the policy of the Department of Education, and no official endorsement by the federal government is intended or should be inferred.

# Contents

	<b>Page</b>
Introduction.....	1
Study Context.....	1
Impact Research Questions.....	2
Logic Model and Literature Base .....	3
Evaluation Logic Model .....	3
Research on Arts Integration .....	5
Research on Professional Development.....	6
Evaluation Methods .....	7
Sample.....	7
Data Collection and Measures .....	9
Analysis Approach.....	13
Findings.....	15
Program Impacts on Teachers’ Use of Arts Integration .....	15
Program Impacts on Students’ Mathematics Achievement.....	16
Discussion.....	17
Recommendations for Scale-Up and Future Research .....	19
References.....	21
Appendix A. Examples of Arts Integration Routines Observed in Classroom Observations .....	A-1
Appendix B. Statistical Models for Impact Analyses.....	B-1
Appendix C. Interpretation of the Size of the Effects of Wolf Trap’s Professional Development Program .....	C-1

# Introduction

## Study Context

In 2010, the [Wolf Trap Foundation for the Performing Arts, Institute for Early Learning Through the Arts](#) (Wolf Trap),<sup>1</sup> was awarded a U.S. Department of Education Arts in Education Model Development and Dissemination (AEMDD) grant ([www2.ed.gov/programs/artsedmodel/index.html](http://www2.ed.gov/programs/artsedmodel/index.html)). The purpose of the AEMDD grant was to develop, implement, and disseminate a research-based program of professional development that equips teachers to infuse mathematics instruction with performing arts strategies in their prekindergarten and kindergarten classrooms. The professional development program developed by Wolf Trap's project team<sup>2</sup> includes annual summer institutes as well as in-classroom coaching during the school year. In the summer institutes, teaching artists in the disciplines of dance, music, and drama work with teachers in teams to develop standards-based performing arts and mathematics experiences. During the school year, teachers and teaching artists work together in partnership, planning and implementing lessons in the classroom—a research-based strategy often called the artist residency model (Burnaford, 2007).

In the first year of the grant, a planning year, Wolf Trap teaching artists explored the conceptual connections between the arts disciplines and the mathematics standards for prekindergarten and kindergarten. They designed lesson plans that described the connections and specified the objectives for both teaching and student learning. During the two years of teaching residencies, teaching artists coached teachers in the classroom as teachers implemented lessons built on the connections between the performing arts and mathematics. With this grant, an initial and a follow-up institute occurred in the summer of the implementation years.

In Wolf Trap artist residencies, artists guided teachers in the process of integrating performing arts strategies by identifying the concept for mathematics skills development and a corresponding performing arts skill and specifying the arts strategy that highlighted the mathematics concepts. Over the period of the residency, the artist first took a more prominent role as the coach and then gradually released the management of the integration to the teacher.

Integration was intended to be consistent with the elements of each performing art and coherent with the mathematics curriculum. For example, one teacher working with a dance artist might integrate mathematics and dance by teaching the mathematical concept of an AB pattern with the dance essential of choreography of movements. Another teacher working with a drama artist

---

<sup>1</sup> Since 1981, the Wolf Trap Institute has served hundreds of thousands of young children ages 3 months old through kindergarten, and their teachers, parents, and caregivers throughout the 17 Wolf Trap Regional Programs ([http://www.wolftrap.org/Education/Institute\\_in\\_your\\_Community/Regional\\_Programs.aspx](http://www.wolftrap.org/Education/Institute_in_your_Community/Regional_Programs.aspx)) and other locations across the country. In the field of early childhood education, the Institute's approach is widely recognized as an effective model for direct instruction of children and embedded professional development for teachers. The Wolf Trap Institute's arts-based teaching method taps into children's innate desire for active, multisensory learning—as children literally embody concepts by singing and dancing—and engages their imagination through puppetry, story dramatization, and role play.

<sup>2</sup> Throughout this report, we refer to Wolf Trap and its project team, which includes the grant director, the associate director for professional development at Wolf Trap's Institute for Early Learning, specialists who routinely plan artist activities in multiple districts, the grant program assistants, and the teaching artists on contract to Wolf Trap for this grant. Early Childhood STEM Learning Through the Arts is the official name of the Wolf Trap grant project.



might use storytelling with pictures, gestures, and sound effects to create a representation of sets of characters in a story, discussing the size and differences among sets. When working with a music artist and studying the value of money, a teacher may introduce differences in amounts with a song and use the musical element of steady beat to highlight the differences in value. The goal in connecting these concepts and essential elements of each discipline is for learning in one subject to enhance, reinforce, or extend learning in the other.

American Institutes for Research (AIR) conducted a four-year evaluation of the Wolf Trap AEMDD grant project beginning in 2010 when the grant was awarded. The purpose of this evaluation was to examine the implementation of the Wolf Trap professional development program and assess its impact on teacher practice (use of performing arts strategies in particular) and students' mathematics knowledge.

An earlier report prepared by AIR (Ludwig & Goff, 2013) presented findings about the implementation of the Wolf Trap AEMDD program, based on the engagement of the first group of three treatment schools and 26 teachers.<sup>3</sup> AIR found that overall, Wolf Trap and the teaching artists delivered professional development that exhibited the six features of high-quality professional development: form, content, active learning, coherence, duration, and collective participation, qualified by the important considerations that (a) not all eligible teachers from each school participated, and (b) not all recruited teachers participated for the entirety of the program. The role of the teaching artist was well defined and implemented as planned with support from content and pedagogy experts.

AIR also found that the Wolf Trap professional development program was implemented with fidelity and delivered preparation to teachers to infuse performing-arts-based strategies into their mathematics instruction, starting in the professional development institutes and then continuing in the residencies. This finding was evidenced in the qualitative data from the institutes, observations, and artist interviews and artifacts: The institutes followed the agenda as planned; the residencies followed the coaching cycle as planned; and the lesson plans were used to meet the goals of content coverage, instruction, and arts integration. Wolf Trap used several approaches to optimize fidelity: a planning year and practice sessions with teaching artists, consistent use of local content experts, and materials structured to reflect the concepts and approaches used in both the institutes and the residencies.

## **Impact Research Questions**

This report focuses on program impacts by addressing two research questions:

1. What was the impact of the Wolf Trap professional development on prekindergarten and kindergarten teachers' use of performing arts and mathematics strategies?
2. To what extent did students in the study treatment schools demonstrate better mathematics knowledge than students in the control schools?

---

<sup>3</sup> An article based on this report will appear in a 2014 special issue of the *Journal for Learning through the Arts*, which is devoted to experiences and lessons learned from a group of AEMDD grantees.

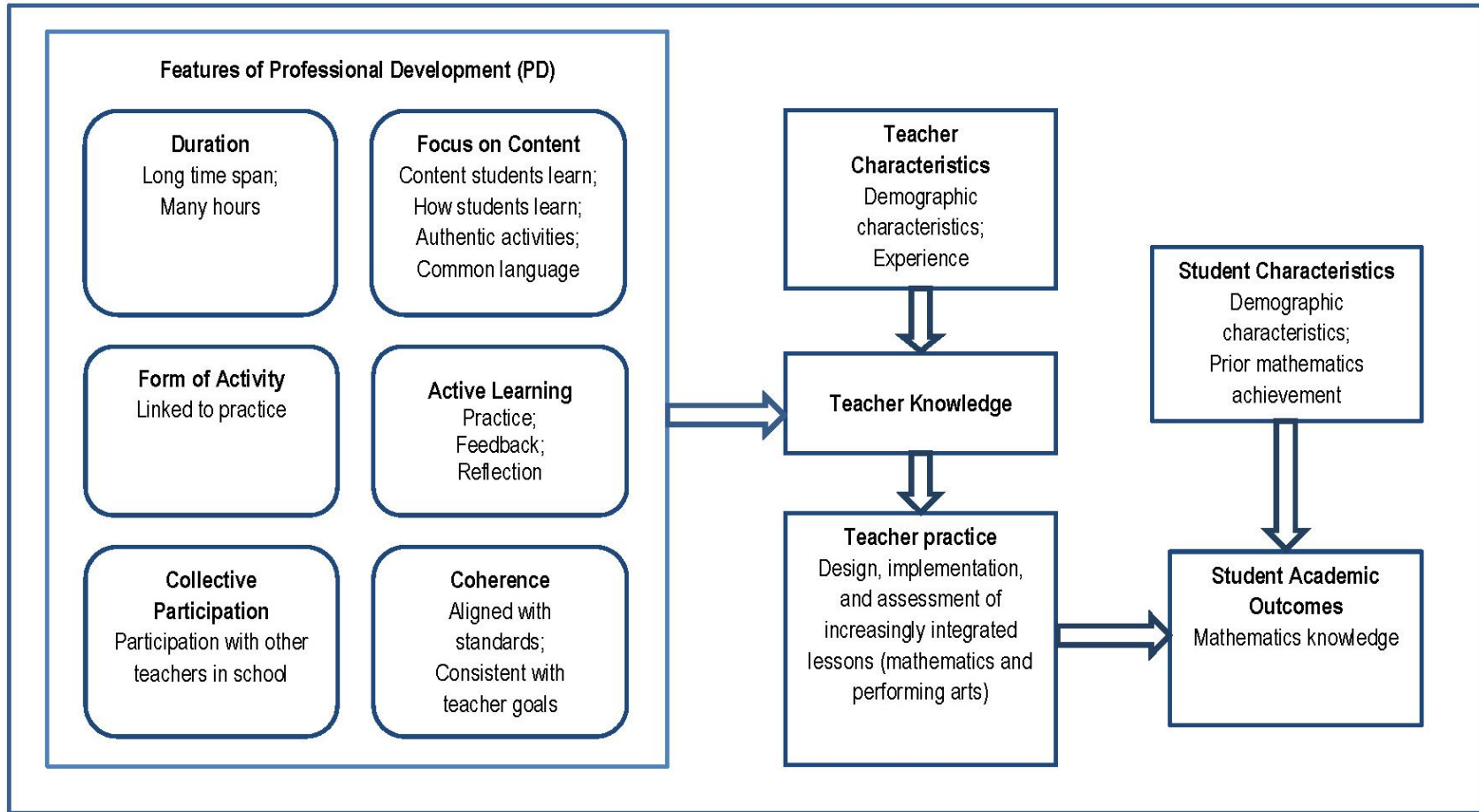
# Logic Model and Literature Base

## Evaluation Logic Model

In designing the evaluation of the AEMDD grant project, AIR developed a conceptual model that described this hypothesis: Professional development with high-quality features would lead to teacher knowledge and teacher practices that would in turn contribute to improved student outcomes (i.e., mathematics knowledge). The relationship between professional development and the teacher practice and student outcomes was based on research on arts integration and research about professional development.

The conceptual model is shown in Exhibit 1 on the following page.

**Exhibit 1. Conceptual Model for the Wolf Trap AEMDD Grant Project Evaluation**



## Research on Arts Integration

The Wolf Trap professional development program evaluated by AIR focused on arts integration, which is defined as instruction that integrates content and skills from the arts with content and skills from other core subjects, with the intent of increasing learning in both areas (Ruppert & Habel, 2011).

Arts integration is often facilitated by teaching artists, who are an essential component of arts-integration programs (Freeman, Seashore, & Werner, 2003). Rabkin and Redmond (2004) define the salient features of arts-integrated instruction as follows:

1. Teacher-artist teams link an art form and an academic discipline.
2. Student groups' work in the art form is central to the experience and to continuous assessment.
3. Content includes material related in meaningful and direct ways to students' experiences.
4. Units have a balanced focus on academic content, academic skills, arts skills, and arts content.
5. Units include basic skills and higher-order skills.
6. Units usually culminate with an artistic product that demonstrates student learning of content and skills and contributes to the public culture of the school community. (p. 137)

Rabkin and Redmond (2004) also note that institutional, school, and community-level elements (e.g., districts' arts standards, current professional development for teachers, and schools' prior experiences with arts) are important to the success and sustainability of arts-integration initiatives.

Research indicates that arts integration has great potential for student learning in multiple disciplines (Burnaford, 2007; Goff & Ludwig, 2013). For example, Ingram and Reidel (2003) reported finding a significant positive link between in-school arts-integrated programming (as part of the Arts for Academic Achievement program) and standardized test scores. Similarly, Catterall and Waldorf (1999) reported that children in Chicago arts-integrated elementary schools performed better on tests than children in control schools. In the field of early childhood education, Erdoğan and Baran (2009) reported that drama-infused math instruction for Turkish six-year-olds was associated with their mathematics achievement test scores. Researchers have also found nonacademic effects in studies of arts integration; for example, teachers and classroom observers have reported improvements in students' creative and critical thinking abilities after arts-integrated programming (Curva et al., 2005; Randi Korn & Associates, 2005; Randi Korn & Associates, 2010).

Research evidence also supports the Wolf Trap approach to arts integration. In a randomized controlled trial of a Wolf Trap-sponsored literacy-focused initiative (Fairfax Pages Professional Development Project), for example, students in treatment schools (schools where teachers participated in the summer professional development and worked with teaching artists in residencies) were found to outperform their counterparts in control schools on the Child

Observation Record scales of initiative, social relations, creative representation, language and literacy, logic and mathematics, and movement and music (Klayman, 2006).

## **Research on Professional Development**

The evaluation of the Wolf Trap professional development model is in the tradition of studies investigating the impact of professional development on teacher practices and student outcomes (Garet et al., 2008; Garet et al., 2011). The professional development features examined in this study are the six features of high-quality professional development identified in prior research: form, focus on content, active learning, coherence, duration, and collective participation (Garet, Porter, Desimone, Birman, & Yoon, 2001). The study also focuses on classroom coaching as an important form of professional development—one that has been shown to lead to teacher acquisition of knowledge and skills (Joyce & Showers, 1995). The literature defines four forms of coaching: technical, problem solving, reflective practice, and building a community of learners (AIR, 2005). The work of the teaching artist can be viewed as technical coaching (i.e., instruction in a particular technique). Teaching artists are focused on teaching behaviors that are applied to a particular subject but that can be applicable more broadly and are closely aligned with the curriculum and pedagogy (Kennedy, 1998). An instructional coach typically supports teachers in the use of instructional strategies in one content area. The teaching artist is required to hold two content areas in balance while understanding the learning needs of early childhood students. This is a unique type of coaching that art educators believe deserves more attention in terms of its impact (Rabkin, Reynolds, Hedberg, & Shelby, 2011).

Over the past decade, additional studies and reviews of studies have offered new insights into how some of the features of high-quality professional development identified in prior research are related to changes in teacher practices and student outcomes. For example, Yoon, Duncan, Lee, Scarloss, and Shapley (2007) reviewed nine rigorous studies of professional development in which the professional development was delivered to teachers directly. They found that studies with professional development programs of more than 14 hours showed a positive and significant effect on student achievement, and that teachers who receive substantial professional development—an average of 49 hours in the nine studies reviewed—markedly boosted student achievement.

However, some researchers are challenging the relationships between those professional development features and teacher practice outcomes. In a recent review of knowledge emerging from studies of professional development, Wilson (2013) notes that rigorous research has “yet to produce conclusive support for those characteristics” and that “problems include a lack of sound measures and [lack of] a strong theoretical understanding of the mechanisms of teacher learning” (p. 311). Some researchers are focusing on defining high-level teacher practices and classroom features associated with student achievement (<http://www.teachingworks.org/>). In the coming years, research on professional development may shift from the general features that we study in this evaluation to professional development for specific practices.

## Evaluation Methods

The evaluation of the Wolf Trap AEMDD grant program was based on a randomized controlled trial in which 22 schools in a large suburban school district were randomly assigned to the treatment and control conditions. In this section, we explain the study sample, measures and data collection, and analytic approach used to address the two research questions.

### Sample

**Study Schools.** For this study, Wolf Trap recruited a total of 22 elementary schools over two years from one large district—six schools in 2011 (Group 1 schools) and 16 schools in 2012 (Group 2 schools). To recruit schools with prekindergarten and kindergarten classrooms for the study, Wolf Trap and its partner school district worked with the superintendent’s office, the research office, and the Title I office of the district to disseminate information about the study in the superintendent’s newsletter, at Title I principal meetings, and in letters to principals. The information in the communications explained the study conditions and requested that teachers from prekindergarten and kindergarten classrooms in each school consider participation. Schools that were willing to participate in the study were randomly assigned to receive the Wolf Trap professional development program (i.e., treatment schools) or the business-as-usual condition (i.e., control schools), with 11 schools in each study condition across the two school groups (three in Group 1 and eight in Group 2).

Group 1 schools implemented the Wolf Trap professional development program during the 2011–12 and 2012–13 school years; Group 2 schools implemented in the program during 2012–13 and 2013–14. All six Group 1 schools remained in the study over the two years of implementation. Four of the 16 Group 2 schools left the study before study activities began, and two additional schools left after the first year of implementation. The treatment schools had an average enrollment of 1,158 students, compared with 1,042 students for the control schools (see Exhibit 2). Among the treatment schools, about half (47.9 percent) of the students on average were eligible for free or reduced-price lunch, 72.6 percent were non-white, and 39.0 percent were English language learners. These percentages were somewhat higher than the percentages for the control schools, but none of the differences were statistically significant ( $p > .05$ ).

### Exhibit 2. School Background Characteristics, by Treatment Status

Characteristics	Treatment Group	Control Group	Difference	P-value
Total school enrollment	1,158.1	1,042.4	116.7	0.519
Percentage of students eligible for free or reduced-price lunch	47.9	39.7	8.2	0.519
Percentage of student who were non-white	72.6	61.6	11.0	0.224
Percentage of students who were English language learners	39.0	37.1	2.0	0.844

*Note.*  $N = 18$  schools (10 treatment and 8 control).

**Teacher Sample.** In each recruited school, all prekindergarten (including Head Start) and kindergarten teachers were invited to participate in the study. Across the two groups of schools, we recruited 80 teachers (48 treatment and 32 control) in total. The number of teachers decreased in each group of schools over the course of the two years of the program because of reasons such as teacher turnover, reassignment of teachers to different grades, and the loss of participating schools. In the second year of implementation, 51 teachers (29 treatment and 22 control) were participating across the two groups of schools.

Of the 80 teachers we recruited, 64 completed a background survey at the beginning of the study that gathered information about their background characteristics and prior experience. As Exhibit 3 shows, among those who returned the background survey, teachers in the treatment schools and teachers in the control schools were generally similar in terms of teaching experience, ethnicity (white versus non-white), primary teaching assignment (prekindergarten versus kindergarten), number of prekindergarten or kindergarten students taught, and whether they had a prekindergarten (PK) or kindergarten (K) certificate. None of the differences between the two groups of teachers were statistically significant ( $p > .05$ ).

**Exhibit 3. Teacher Background Characteristics, by Treatment Status**

Characteristics	Treatment Group	Control Group	Difference	P-value
Years of teaching experience (percent)				
3 years or fewer	24.3	28.2	-3.9	0.752
4–10 years	32.4	21.9	10.6	0.352
More than 10 years	43.2	49.5	-6.2	0.637
Non-white (percent)	16.7	11.5	5.2	0.572
Teaching assignment: PK (percent)	48.6	38.6	10.0	0.613
Number of PK/K students taught	18.7	19.6	-0.9	0.676
PK/K certificate (percent)	75.7	88.5	-12.9	0.250

*Note.* Sample size = 18 schools, 61–64 teachers.

**Student Sample.** From the classroom of each lead participating teacher,<sup>4</sup> we randomly selected eight students for the assessment of mathematics knowledge, the key student outcome measure. These students, with parental consent, were assessed twice in the first year of their school’s participation (baseline and spring) and once more in the spring of the second year of participation. If students remained in the district but were in a different school, the test examiners went to the school for the last follow-up assessment administration. Assessment was voluntary, so that if students’ parents changed their mind about consent, students were not assessed. There were also a few cases in which principals in the new receiving school did not wish to participate in this activity.

<sup>4</sup> The lead teacher is the teacher assigned the instructional responsibility for the classroom. Prekindergarten and kindergarten teachers in the study also had instructional aides. In study schools, special educators may have worked with the lead teachers.

Across both groups of schools, 421 students (247 treatment and 174 control) were assessed at baseline, 394 students (220 treatment and 174 control) were assessed in the spring of the first year of implementation, and 354 students (198 treatment and 156 control) were assessed in the spring of the second year of implementation.<sup>5</sup> Students in the two study conditions had similar background characteristics (i.e., baseline mathematics assessment scores, ethnicity, and gender), as shown in Exhibit 4.

#### Exhibit 4. Student Background Characteristics, by Treatment Status

Characteristics	Treatment Group	Control Group	Difference	P-value
Baseline mathematics assessment score	11.0	10.2	0.7	0.819
Non-white (percent)	75.2	63.2	12.0	0.444
Female (percent)	44.6	42.5	2.1	0.665

*Note.* Sample size = 18 schools, 421–441 students.

#### Data Collection and Measures

AIR’s evaluation of the Wolf Trap AEMDD project draws on multiple data sources from teachers, teaching artists, and students. Exhibit 5 shows the data collection activities, participants, and schedule for the overall evaluation. A description of the data collection instruments follows this exhibit.

#### Exhibit 5. Data Collection and Participants

Data Collection Activity	Intended Participants	Schedule of Data Collection
Teacher baseline survey <sup>a</sup>	All treatment and control teachers	Group 1: August 2011 (treatment), September 2011 (control) Group 2: August 2012 (treatment), September 2012 (control)
Observations of classrooms <sup>b</sup>	All treatment and control teachers	Group 1 Teachers: January 2012, May/June 2012; October/November 2012, May/June 2013 Group 2 Teachers: October/November 2012, May/June 2013; October/November 2013, May/June 2014
Online survey of teachers regarding Wolf Trap services	All treatment teachers	Spring 2011, 2012, 2013 conducted by Wolf Trap
Observations of the summer professional development institutes <sup>c</sup>	Treatment teachers and teaching artists	August 2011, 2012, 2013 (5 days first institute, 3 days follow-up for each group of treatment teachers)

<sup>5</sup> The first year of implementation is 2011–12 for Group 1 schools and 2012–13 for Group 2 schools; the second year of implementation is 2012–13 for Group 1 schools and 2013–14 for Group 2 schools.



Data Collection Activity	Intended Participants	Schedule of Data Collection
Teaching artist residency planning forms <sup>d</sup> and lesson plan forms <sup>e</sup>	Teaching artists and treatment teachers in residency activities	Submitted to AIR June/July 2013; 2014 at the end Group 1 and Group 2 residencies
Interviews with teaching artists <sup>f</sup>	All teaching artists (9)	June 2013
Early Math Diagnostic Assessment (EMDA), an instrument developed by Pearson and distributed by PsychCorp	Sample of 8 students from the class taught by each treatment and control teacher	Group 1: Fall 2011 (baseline); Spring 2012; Spring 2013 Group 2: Fall 2012 (baseline); Spring 2013; Spring 2014

<sup>a</sup> Survey available at

[http://www.wolftrap.org/~media/files/pdf/education/wolf\\_trap\\_teacher\\_survey.ashx](http://www.wolftrap.org/~media/files/pdf/education/wolf_trap_teacher_survey.ashx)

<sup>b</sup> Classroom observation form available at

[http://www.wolftrap.org/~media/files/pdf/education/classroom\\_observation\\_form.ashx](http://www.wolftrap.org/~media/files/pdf/education/classroom_observation_form.ashx)

<sup>c</sup> Professional Development Institute observation form available at

[http://www.wolftrap.org/~media/files/pdf/education/blank\\_2012\\_pd\\_obsform.ashx](http://www.wolftrap.org/~media/files/pdf/education/blank_2012_pd_obsform.ashx)

<sup>d</sup> Example residency planning form available at

[http://www.wolftrap.org/~media/files/pdf/education/Planning\\_Form.ashx](http://www.wolftrap.org/~media/files/pdf/education/Planning_Form.ashx)

Residency planning form analysis rubric available at

[http://www.wolftrap.org/~media/files/pdf/education/blank\\_residency\\_planning\\_form\\_analysis\\_spreadsheet.ashx](http://www.wolftrap.org/~media/files/pdf/education/blank_residency_planning_form_analysis_spreadsheet.ashx)

<sup>e</sup> Example lesson plan available at

[http://www.wolftrap.org/~media/files/pdf/education/Early\\_STEM\\_Arts\\_Lesson\\_Plan.ashx](http://www.wolftrap.org/~media/files/pdf/education/Early_STEM_Arts_Lesson_Plan.ashx)

Lesson plan analysis spreadsheet available at

[http://www.wolftrap.org/~media/files/pdf/education/blank\\_lesson\\_plan\\_analysis\\_spreadsheet.ashx](http://www.wolftrap.org/~media/files/pdf/education/blank_lesson_plan_analysis_spreadsheet.ashx)

<sup>f</sup> Interview protocol available at

[http://www.wolftrap.org/~media/files/pdf/education/Interview\\_protocol.ashx](http://www.wolftrap.org/~media/files/pdf/education/Interview_protocol.ashx)

Information about the data collection activities was first disseminated to principals and teachers through introductory letters about the project. At key points in the project schedule, information was prepared and shared with teachers and parents about the type, schedule, and purpose of data collection: at entry into the study, at survey administration, prior to classroom observations, and prior to sampling students for participation in the EMDA assessment.

**Surveys.** AIR administered the Professional Development and Instructional Practice (PDIP) survey to both treatment and control teachers. The PDIP survey contains items about the teacher's assigned grade level, number of students, teaching experience, professional development experience, and instructional practices in teaching mathematics.<sup>6</sup> The items have been analyzed and shown to be reliable and valid. Wolf Trap also conducted an online services-focused survey with treatment teachers. A total of 64 teachers responded to the background surveys from both groups.

**Observation Forms.** Two observation forms were developed for this study. The observation form for the summer institutes was based on professional development observation forms used in two national studies of professional development impact on reading and mathematics (Garet et al., 2008; Garet et al., 2011). This form was grounded in the agenda for the summer institutes

<sup>6</sup> The PDIP survey has been used in prior studies conducted by AIR, which examined the professional development delivered by Math and Science Partnership Projects funded by the National Science Foundation.

and the professional development elements that were anticipated (e.g., a focus on mathematics content, inclusion of arts, and linking of arts and mathematics). After researchers attended the first summer institute, a framework emerged from the materials, training, and resources for participants. The classroom observation form was developed on the basis of this framework. When planning for classroom observations, we requested permission to observe for 45 minutes when mathematics was being taught without specifying that the teachers should be using the strategies they had been implementing with the teaching artist.

In the literature, there is a lack of standardized instruments to measure arts integration. For this study, we created an observation form drawing on the constructs in the Wolf Trap model framework, which reflects the key features of arts integration articulated by Rabkin and Redmond (2004). The four main measures of arts integration and their descriptions in the observation form are provided in Exhibit 6.

**Exhibit 6. Measure of Arts Integration**

Measures of Arts Integration	Description in Observation Form
Linking Arts With Math	Today’s teacher linked an art form with math study through common concepts, language, or activities, reflecting the professional development provided.
Student Group Work in Art	Today’s teacher ensured that student group work in the performing art form was central to the experience. Materials were provided, and the teacher explained the purpose of using the materials.
Balanced Focus	Today’s unit(s) balanced focus on academic content and skills with arts content and skills.
Art Product	Today’s unit(s) culminated in an artistic product, demonstrating student learning of academic content and skills.

Based on what they observed in the class, the observers coded each of the four key features as *Not Evident*, *Partially Implemented*, or *Fully Implemented*, with a score of 0, 1, and 2 respectively, based on the rubric shown in Exhibit 7 (see Appendix A for examples).

### Exhibit 7. Rubric for Measuring Features of Arts Integration

Feature	Not Evident	Partially Implemented	Fully Implemented
Linking Arts With Math	Concepts, language, or activities related to the performing arts area or to the math topic were a focus of the lesson, but the teacher did not explicitly link the disciplines.	Common concepts, language, or activities were used for the purpose of linking the performing arts and math, but the link between the two was not fully implemented (e.g., space in dance was a vehicle for a math lesson, but it was not clearly linked to the related math concept of number, shapes, or position).	Common concepts, language, or activities were used for the purpose of linking the performing arts and math study, and the implementation was faithful to the examples or elaborations of teaching artist lessons (e.g., the teacher used dance to demonstrate shapes, rhythm to demonstrate patterns, and/or aspects of a story to demonstrate counting).
Student Group Work in Art	The students did not do group work. The teacher provided materials for student activities but provided little setup or explanation, so students did not use the materials as planned.	The students engaged in group work, and the teacher provided materials essential to the group activity in the art form and an explanation about their use but did not indicate the use of materials, and the engagement with the art form was not fully central to the experience.	The students engaged in group work. The teacher provided materials essential to the group activity in the art form. The teacher provided an explanation about the use of the materials, and the planned lesson, in a group, was fully implemented, so the work in the performing arts was central to the experience.
Balanced Focus	Only the performing arts or math content was present in the lesson (e.g., the teacher performed a math lesson without using the arts).	Both the performing arts and math content were present in the lesson; however, the balance was uneven (e.g., it was difficult to determine the focus of the lesson).	Both the performing arts and math content were present in the lesson, and there was a balance that reflected the approach taught in the Wolf Trap professional development.
Art Product	The lesson ended without a clear artistic product or activity. No opportunity for students to demonstrate their math or art learning was provided.	The lesson ended with an artistic product or activity; however, it was a product or activity that did not provide a full opportunity for students to demonstrate math learned.	Activities led to a fully developed song, dance, or dramatic presentation, in which students were able to demonstrate their learning of the related math content (e.g., a fully developed song that reinforced counting, a fully developed dance that reinforced shapes).

**Documentation From Artists.** At the end of each set of teaching residencies, AIR received documentation from Wolf Trap submitted by the teaching artists that included the residency forms and the lesson planning forms. The residency form was the overall plan for coaching. It identified the curriculum standard for the residency and other skills (social-emotional, language, cognitive, and motor skills), goals (math standard, curriculum focus, arts-strategy skill), and schedule. The plan was accompanied by a debriefing form with stimulus questions to process the outcomes with the teacher. The individual lesson plans are intentionally structured for coherence with the residency form and cover arts and mathematics concepts, professional development skill focus, identification of vocabulary, objectives, teacher questions, procedure, assessment strategies, modifications, extension experiences, and a description of the lesson procedure.

**Interviews.** AIR interviewed the nine teaching artists working with the first group of treatment teachers in 2011–13. These hour-long phone interviews asked the artists about their overall experience in the role, their preparation for the Wolf Trap project, and their views about the experience of the teachers with whom they worked.

**Student Assessment.** The Early Math Diagnostic Assessment (EMDA) was selected as the measure of mathematics knowledge for this evaluation, which is aligned with curriculum standards of the participating school district. The selection of the EMDA was made by the partner district and the Department’s technical program officer.<sup>7</sup> EMDA is a norm-referenced instrument, individually administered, and appropriate for learners in Grades PK–3. “The EMDA tests (Math Reasoning and Numerical Operations) are derived from the previously published Wechsler Individual Achievement Test, Second Edition (WIAT-II, 2001)” (Pearson Clinical, 2003, p. 1). This study focused on the EMDA math reasoning test, which includes 45 items testing children’s math reasoning knowledge and skills, including counting, ordering numbers, identifying and comparing shapes, creating and solving problems using whole numbers, using patterns to solve problems, telling time and using time to compare and order events, solving problems using or related to money, using grids and graphs, using non-standard and standard units of measure.

The project team hired test examiners who were former teachers, aides, and teaching artists who were familiar with working in a school setting. AIR trained the test examiners before baseline assessment began and repeated training as new test examiners were hired. AIR also conducted refresher training in spring 2014, prior to the test administration. The same test was used in each administration. The examiner’s manual provides tables that translate the raw scores to level of proficiency specific to grade and the time of the year (fall, spring) of administration.

## **Analysis Approach**

In this section, we describe our analysis approach to addressing the research questions about the impacts of the Wolf Trap professional development program on teachers’ use of both performing arts and mathematics strategies (Research Question 1) and students’ mathematics knowledge (Research Question 2).

---

<sup>7</sup> The evaluation design originally called for using a district assessment, the Kindergarten Mathematical Reasoning Assessment (K-MRA) and a similar assessment for Grade 1; both are formative assessments supported by a district online platform.

## **Analysis of Impact on Teacher Practice**

The analysis of the impacts of the Wolf Trap professional development program on the extent to which teachers integrated specific arts approaches within their mathematics lessons was based primarily on the four waves of classroom observation data. For each lesson observed, we judged whether it provided opportunities for arts integration, and if it did, we further rated the extent of implementation of the four features of arts integration (i.e., Linking Arts With Math, Student Group Work in Art, Balanced Focus, and Art Product).

We hypothesized that treatment teachers' lessons are more likely to provide opportunities for arts integration, and we tested this hypothesis using a logistic regression model that predicts the probability that a lesson offered such opportunities with treatment status and teacher background characteristics (e.g., teaching experience, ethnicity (white versus non-white), and teaching assignment (PK versus K)).<sup>8</sup> The analysis was conducted for each of the four waves of observations separately. Appendix B provides details about the model.

Because the analyses of the program's impact on arts integration were restricted to lessons that provided opportunities for arts integration, the analysis sample size was much smaller than the number of lessons observed at each observation wave. To maximize the analysis sample size, we created measures of arts integration by pooling data across the four waves and computed the average score of each arts-integration measure (i.e., Linking Arts With Math, Student Group Work in Art, Balanced Focus, and Art Product) across all lessons observed for each teacher.<sup>9</sup> We also computed an overall score of arts integration as the average across all four measures and observation waves. The program's impacts on these five measures of arts integration were assessed using a two-level hierarchical linear model (HLM) that takes into account the clustering of teachers within schools. Details about the model are provided in Appendix B.

## **Analysis of Impact on Student Achievement**

The assessment of the impacts of the Wolf Trap professional development program on students' mathematics achievement was also based on HLM analyses, which take into account the clustering of students within schools. Using EMDA data pooled across both Group 1 and Group 2 schools, we assessed both the one-year impact and the two-year impact of the program on students' mathematics achievement (see Appendix B for details about the model).

---

<sup>8</sup> Ideally, this analysis should be conducted using a multilevel logit model that takes into account the clustering of teachers within schools. However, our analysis sample size is too small for the multilevel logit model given the substantial amount of missing data due to teacher attrition and scheduling problems.

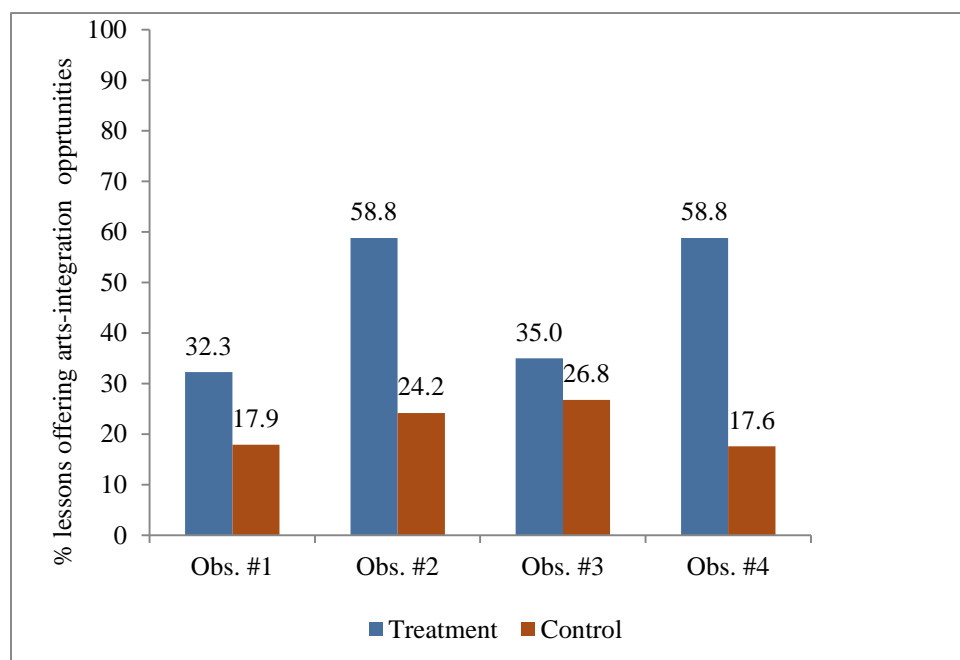
<sup>9</sup> The average scores thus computed may be based on different numbers of lessons for different teachers, depending on how many times a teacher was observed.

## Findings

### Program Impacts on Teachers' Use of Arts Integration

To address our first research question, we assessed the impacts of the Wolf Trap professional development program on both the probability that a lesson provided opportunities for arts integration and the five measures of arts integration. As Exhibit 8 shows, about one third (32.3 percent) of the observed lessons taught by treatment teachers provided opportunities for arts integration, compared with 17.9 percent for control teachers at the first wave of observation. Treatment teachers' lessons also were more likely to provide such opportunities than were control teachers' lessons at the other three waves of observation. None of the differences, however, were statistically significant at the .05 level, most likely because of the small sample size.

**Exhibit 8. Percentage of Observed Classes Providing Arts-Integration Opportunities at Each Wave of Observation, by Treatment Status**



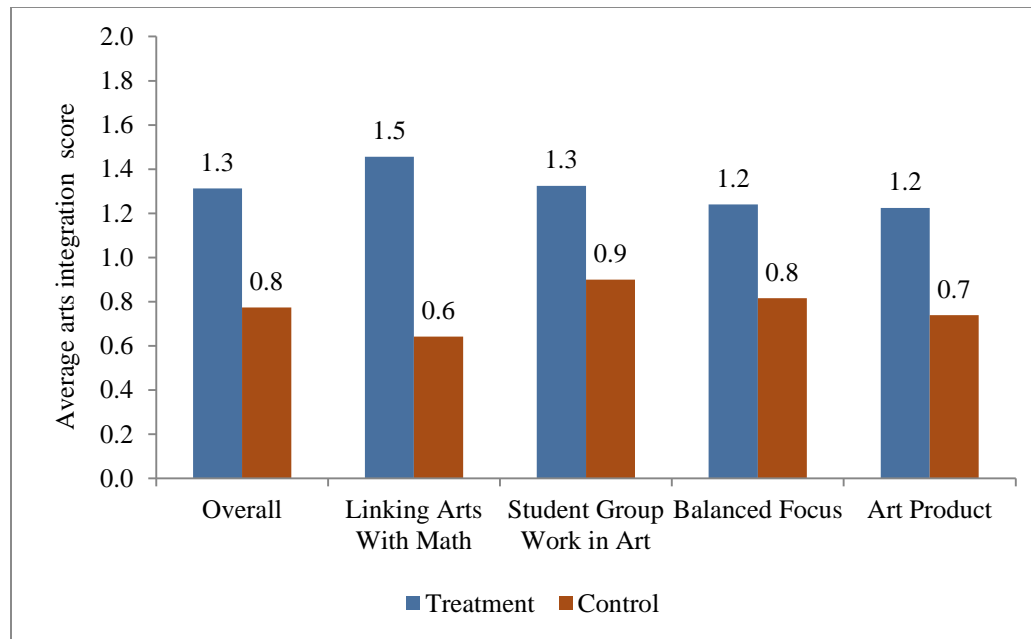
*Notes.*  $N = 47, 24, 21,$  and  $19$  teachers for the four waves of observations, respectively.

The percentage for the treatment group for each measure of arts integration is the unadjusted treatment group mean; the percentage for the control group for each measure was computed on the basis of the treatment group percentage and the estimated group difference.

Among those lessons that offered opportunities for arts integration, the lessons taught by treatment teachers had a higher score than the lessons taught by control teachers on both the overall measure and the four individual measures of arts integration. As Exhibit 9 shows, for example, the average score of the overall measure of arts integration across all lessons observed is 1.3 for treatment teachers, which is significantly higher than the score (0.8) for control teachers ( $p < .05$ ). The difference between treatment and control teachers also was statistically significant for the measure of Linking Arts With Math (1.5 for treatment versus 0.6 for control).

Differences between the two groups of teachers in the other three measures, although in the expected direction, were not statistically significant ( $p > .05$ ), which may be due to the limited sample size.

**Exhibit 9. Average Level of Arts Integration Across Observation Windows, by Treatment Status**



Notes. Sample size = 31–32 teachers; 12–13 schools.

All measures of arts integration are on a 0–2 scale, with 0 representing *Not Evident*, 1 representing *Partially Implemented*, and 2 representing *Fully Implemented*.

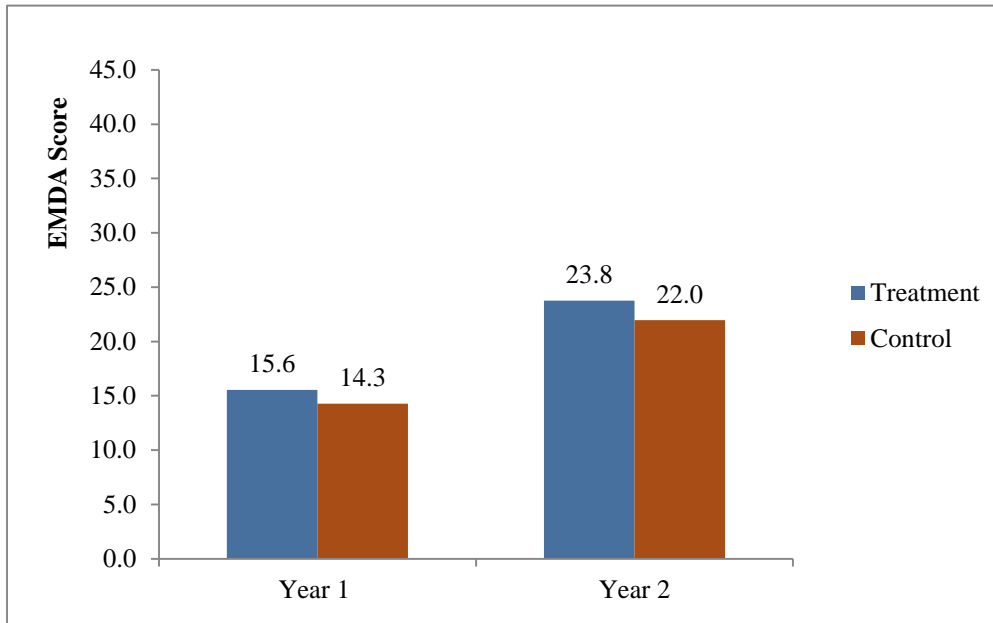
The score for the treatment group for each measure of arts integration is the unadjusted average score; the score for the control group for each measure was computed by subtracting the estimated group difference from the unadjusted average score for the treatment group.

**Program Impacts on Students’ Mathematics Achievement**

Results from our analyses indicate that the Wolf Trap professional development program had a statistically significant positive impact on students’ mathematics achievement in both the first and the second year of implementation. As Exhibit 10 shows, students in the treatment schools outperformed their peers in the control schools by 1.3 points on average on EMDA in Year 1 (15.6 for treatment as compared with 14.3 for control). The difference is statistically significant ( $p < .01$ ) with an effect size of 0.17.<sup>10</sup> The positive impact of the Wolf Trap professional development program persisted in Year 2, with treatment students outperforming the control students by 1.8 points on EMDA ( $p < .01$ , effect size = 0.21). These effect sizes translate to 1.3 and 1.7 additional months of learning, respectively (see Appendix C for further details about the interpretation of the effects).

<sup>10</sup> Effect size is a measure of the magnitude of the treatment effect in standard deviation units. It was computed as the difference in the average EMDA score between the two study groups divided by the pooled within-group standard deviation of the scores.

### Exhibit 10. Average EMDA Math Reasoning Scores in Spring Year 1 and Spring Year 2, by Treatment Status



*Notes.*  $N = 369$  students and 18 schools for Year 1 analysis; 334 students and 18 schools for Year 2 analysis. The average EMDA score for the treatment group in each year is the unadjusted average score; the average score for the control group was computed by subtracting the estimated group difference from the unadjusted average score for the treatment group. The EMDA contains 45 items in total on the math reasoning test for students in prekindergarten through first grade.

## Discussion

Based on a randomized controlled trial, this study assessed the impacts of the Wolf Trap professional development program on both teachers' use of performing arts strategies integrated with mathematics content and the mathematics knowledge of students in prekindergarten and kindergarten. Drawing on classroom observation data, we found that treatment teachers' lessons were more likely to provide opportunities for arts integration than control teachers' lessons at each of the four waves of observation. The differences between the two groups, however, were not statistically significant at the .05 level, possibly because of the small sample size.

Among the observed lessons that offered opportunities for arts integration, the lessons taught by treatment teachers demonstrated a significantly higher level of arts integration than the lessons taught by control teachers on both the overall measure of arts integration and one of the four specific measures of arts integration (i.e., Linking Arts With Math). The impacts of the professional development program on the other three measures of arts integration, although in the desired direction, were not statistically significant. Our analyses also suggest that the Wolf Trap professional development program had a statistically significant positive impact on students' mathematics knowledge in both years of implementation (effect size = 0.17 for the first year and 0.21 for the second year).



These findings should be interpreted with caution given the limitations of this study. First, the sample for the study was relatively small, which may have limited the statistical power of the study to detect true program impacts. Second, school and teacher participation decreased over time because of reasons such as reassignment of teachers to different grade levels and teacher turnover. Attrition, particularly differential attrition between the treatment and control groups, may have introduced bias to the estimated impacts of the professional development program. Third, this study focused on teacher practice and did not measure teacher knowledge, which may have been affected by the professional development intervention as well. Finally, the study only examined one student achievement outcome and did not measure student competencies other than mathematics knowledge that may have been developing as a result of the professional development program. Measures of those additional teacher and student outcomes would be helpful for both the program developer and evaluators.

Despite the limitations, this study produced promising evidence for the impacts of the Wolf Trap professional development program on both teaching and learning. The following factors may have contributed to these positive results:

- **Treating the Earliest Grade Level.** Wolf Trap worked with prekindergarten (including Head Start) and kindergarten classes. For many students, this class was their first introduction to the formal school environment and the first opportunity to learn English. Teachers commented that the use of music, movement, and dramatizing concepts was beneficial for all students, but in particular students who were shy, who had never been to school, or who were speaking another language.
- **Contribution to Teachers' Mathematics Instruction.** Teachers were expected to know the district standards, and most had many years of classroom teaching experience. The use of performing arts strategies linked to mathematics concepts may not have changed teachers' knowledge of mathematics, but it may have provided an instructional boost—making the abstract concepts seem more real and accessible for the class through the new strategies applied by the teacher and the teaching artist.
- **Teacher Enthusiasm for the Arts in the Classroom.** One would expect that differences between schools in the two conditions would not be significant given the level of standardization across schools in the mathematics standards and instructional approaches already used; however, it may be that the teachers in the treatment schools were highly receptive to the professional development and eager to implement new strategies that they believed would result in the improvement of mathematics performance.
- **Improved Classroom Interaction.** Wolf Trap summer institutes and residencies introduced strategies such as giving children the opportunity to lead, asking intentional questions, and giving formative feedback. The increased attention given to student participation, teacher feedback, and improved classroom structure may have contributed to student learning.
- **Benefits of Residencies.** The participation of teachers in the summer institutes varied. AIR and Wolf Trap documented the attendance of treatment teachers at the professional development institute sessions and calculated the number of treatment teachers who participated in at least 75 percent of these sessions and the number who participated in two full years of artist residencies, which meant that each teacher had worked with two

artists. In Group 1, of the 14 treatment teachers who participated in two full years of the study, nine participated in 75 percent or more of the professional development institute sessions. In Group 2, of the 15 treatment teachers who participated in two full years of residencies, seven participated in 75 percent or more of the professional development institute sessions. Teaching artists suggested that the real work around arts integration occurred in the residencies, which may be even more beneficial than the summer institutes for teachers to learn how to incorporate arts strategies into their classes.

- **Variation in Implementation.** Wolf Trap has been developing its combined professional development of institutes and residencies over some time and has done extensive replication through its affiliate network. An operating principle of this model is an expectation that teaching artists will work with teachers as a team in developing strategies and activities that support curriculum in place, classroom conditions, and teachers' objectives and variations in implementation are expected and accepted. Weinbaum and Supovitz (2010) report that there is no definitive evidence indicating whether interventions that are more specific are better than ones with leeway for adaptation. Researchers suggest considering four important areas when implementing a standard intervention: what developers choose to emphasize, the level of complexity of the changes they are expecting, how they engage teachers and administrators in the change process, and the ongoing support for change (Schiffman, Riggan, Massel, Goldwasser, & Andersen, 2008). The Wolf Trap professional development appears successful in its support for adaptation.

The Wolf Trap AEMDD professional development intervention encountered many of the typical challenges faced by professional development developers and implementers, such as school attrition, teacher turnover, student mobility, subject standards revisions, and inconsistent participation in summer institutes. However, findings from this study indicate that when arts strategies and activities were integrated into the teaching of mathematics in prekindergarten and kindergarten classrooms, teachers who were coached to use the strategies were able to apply those strategies independently in their classrooms, and their students performed better than their peers on mathematics tasks consistent with their grade-level standards.

## Recommendations for Scale-Up and Future Research

When research indicates an intervention is promising, developers are likely to consider the next stage: how to expand the intervention from a limited number of sites to a larger number of sites with diverse characteristics so that the program can have a broader impact—i.e., scaling up. In fact, Wolf Trap has been supporting dissemination of its professional development model by holding institutes for all of its 17 affiliates, extending technical assistance through affiliate visits, and training master teaching artists. Each affiliate implements the practices learned through the Wolf Trap training and materials according to its setting, schools served, and financial supports. The positive impacts of the Wolf Trap program on students found in this study as well as a prior study (Klayman, 2006) provide research-based support for the replication of the Wolf Trap professional development model on a larger scale.

Based on the evaluation conducted for the AEMDD grant, we offer the following recommendations for future research:

**Validation of a Rubric Measuring the Practice of Arts Integration.** In a review of assessment approaches used in arts-integrated settings, we learned there are few non-program-specific frameworks or rubrics that would (1) pinpoint measurable aspects of quality arts-integrated instruction and (2) offer a measurement of the overall quality of a given interval of a teacher's arts-integrated instruction. In this AEMDD evaluation, we used a rubric to measure four features of arts integration. The Wolf Trap affiliate programs have the potential to conduct validation studies of this rubric. Further, after the rubric has been validated, it could inform the design of the professional development program for teachers and teaching artists based on a standard of arts-integration classroom practice.

**Measuring Arts Learning.** In this evaluation, our focus was on the measurement of academic knowledge students learned through mathematics instruction integrated with performing arts strategies and activities. We did not seek to measure students' knowledge and skills in the arts disciplines that were likely developing among students in schools participating in the interventions. A recent review conducted by the National Endowment for the Arts on the status of assessment in the arts indicates more research and development are needed (Herpin, Washington, & Li, 2012). We believe that the classrooms, teachers, and students who are working with Wolf Trap would provide an excellent setting for developing such measures and investigating what children are learning in terms of the vocabulary and the essentials of arts disciplines, and further, exploring arts-related outcomes in later years as students have opportunities to study arts disciplines in electives or required courses.

**Learning From Variation in Participation in Summer Institutes versus Artist Residencies.** We learned from interviews with the teaching artists that the artist residencies in general were more likely to influence teacher practice than summer institutes where training also occurred. We also noted that teachers who persisted in the study engaged in both planned residencies. Therefore, the residencies may have been a key to the positive outcomes for students taught by the treatment teachers. We recommend that Wolf Trap explore ways to improve the effects of the summer institutes on teachers. It may be helpful to have briefer institutes with a stronger focus on the activities, terminology, and expectations of the residencies.

## References

- American Institutes for Research. (2005). *Conceptual overview: Coaching in the professional development impact study*. Unpublished manuscript.
- Bloom, H. S., Hill, C. J., Black, A. B., & Lipsey, M. W. (2008). Performance trajectories and performance gaps as achievement effect-size benchmarks for educational interventions. *Journal of Research on Educational Effectiveness*, 1(4), 289–328.
- Burnafor, G. (with Brown, S., Doherty, J., & McLaughlin, H. J.). (2007). *Arts integration frameworks, research and practice: A literature review*. Washington, DC: Arts Education Partnership. Retrieved from <http://choice.dadeschools.net/rrm/resources/BurnaforArtsIntegrationFrameworksResearchPracticeALitReview.pdf>
- Catterall, J. S., & Waldorf, L. (1999). Chicago Arts Partnerships in Education: Summary evaluation. In E. B. Fiske (Ed.), *Champions of change: The impact of the arts on learning* (pp. 47–62). Washington, DC: Arts Education Partnership. Retrieved from <http://artsedge.kennedy-center.org/champions/pdfs/ChampsReport.pdf>
- Curva, F., Milton, S., Wood, S., Palmer, D., Nahamia, C., Radcliffe, B., et al. (2005). *Artful Citizenship Project: Three-year project report*. Tallahassee, FL: Curva and Associates. Retrieved from [http://www.artfulcitizenship.org/pdf/full\\_report.pdf](http://www.artfulcitizenship.org/pdf/full_report.pdf)
- Erdoğan, S., & Baran, G. (2009). A study on the effect of mathematics teaching provided through drama on the mathematics ability of six-year-old children. *Eurasia Journal of Mathematics, Science & Technology Education* 5(1), 79–85. Retrieved from [http://www.ejmste.com/v5n1/EURASIA\\_v5v1\\_SErdogan.pdf](http://www.ejmste.com/v5n1/EURASIA_v5v1_SErdogan.pdf)
- Freeman, C., Seashore, K. R., & Werner, L. (2003). *Arts for academic achievement: Models of implementing arts for academic achievement: Challenging contemporary classroom practice*. Minneapolis, MN: University of Minnesota, Center for Applied Research and Educational Improvement.
- Garet, M. S., Cronen, S., Eaton, M., Kurki, A., Ludwig, M., Jones, W., et al. (2008). *The impact of two professional development interventions on early reading instruction and achievement*. Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance.
- Garet, M., Porter, A., Desimone, L., Birman, B., & Yoon, K. S. (2001). What makes professional development effective? Results from a national sample of teachers. *American Educational Research Journal*, 38(4), 915–945.
- Garet, M., Wayne, A., Stancavage, F., Taylor, J., Eaton, M., Walters, K., et al. (2011). *Middle school mathematics professional development impact study: Findings after the second year of implementation* (NCEE 2011-4024). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance.

- Goff, R., & Ludwig, M. (2013). *Teacher practice and student outcomes in arts-integrated learning settings: A review of literature*. Washington, DC: American Institutes for Research. Retrieved from <http://www.air.org/resource/teacher-practice-and-student-outcomes-arts-integrated-learning-settings-review-literature>
- Herpin, S. A., Washington, A. Q., & Li, J. (2012). *Improving the assessment of student learning in the arts: State of the field and recommendations*. Washington, DC: National Endowment for the Arts. Retrieved from <http://arts.gov/sites/default/files/WestEd.pdf>
- Ingram, D., & Riedel, E. (2003). *What does arts integration do for students?* Minneapolis, MN: University of Minnesota, Center for Applied Research and Educational Improvement.
- Joyce, B., & Showers, B. (1995). *Student achievement through staff development*. White Plains, NY: Longman.
- Kennedy, M. (1998). *Form and substance in inservice teacher education* (Research Monograph No. 13). Madison, WI: National Institute for Science Education, University of Wisconsin–Madison. Retrieved from <http://www.losmedanos.edu/deved/documents/KennedyeffectsofPD.pdf>
- Klayman, D. (2006). *Executive summary of the final evaluation report for Fairfax Pages Professional Development Project: An effective strategy for improving school readiness*. Potomac, MD: Social Dynamics.
- Ludwig, M., & Goff, R. (2013). *The implementation of the Wolf Trap Early Childhood STEM Learning Through the Arts AEMDD grant project: Interim evaluation report*. Retrieved from [http://www.wolftrap.org/~media/files/pdf/education/stem\\_interim\\_evalreport.pdf?la=en](http://www.wolftrap.org/~media/files/pdf/education/stem_interim_evalreport.pdf?la=en)
- Ludwig, M., & Song, M. (2015). *Evaluation of professional development in the use of arts-integrated activities with mathematics content: Findings from the Evaluation of the Wolf Trap Arts in Education Model Development and Dissemination Grant*. Washington, DC: American Institutes for Research.
- Ludwig, M., Song, M., Kouyate-Tate, A. F., Cooper, J., Phillips, L., & Greenbaum, S. (2014). Evaluation of professional development in the use of arts-integrated activities with mathematics content: Findings about program implementation. *Journal for Learning through the Arts*, 10, 1. Retrieved from <https://escholarship.org/uc/item/1905c5tm>
- Pearson Clinical. (2003). *Early math diagnostic assessment examiner's manual*. Available from <http://www.pearsonclinical.com/education/products/100000208/early-math-diagnostic-assessment-emma.html>
- Rabkin, N., & Redmond, R. (2004). *Putting the arts in the picture: Reframing education in the 21st century*. Chicago, IL: Columbia College.

- Rabkin, N., Reynolds, M., Hedberg, E., & Shelby, J. (2011). *A report on the Teaching Artist Research Project: Teaching artists and the future of education*. Chicago, IL: National Opinion Research Center at the University of Chicago. Retrieved from [http://www.teachingartists.com/images/RabkinN\\_Teach\\_Artist\\_Research\\_2011.pdf](http://www.teachingartists.com/images/RabkinN_Teach_Artist_Research_2011.pdf)
- Randi Korn & Associates. (2010). *Educational research: The art of problem solving*. Alexandria, VA: Museum Visitor Studies, Evaluation and Audience Research. Retrieved from [http://www.guggenheim.org/images/lta/pdfs/APS\\_Executive\\_Summary.pdf](http://www.guggenheim.org/images/lta/pdfs/APS_Executive_Summary.pdf)
- Randi Korn & Associates. (2005). *Solomon R. Guggenheim Museum: Teaching literacy through arts. Year One: 2004–2005 study*. Alexandria, VA: Museum Visitor Studies, Evaluation and Audience Research.
- Ruppert, S., & Habel, S. (2011, July 19). Integrating the arts across the curriculum [webinar]. *Education Week Webinar*. Retrieved from <http://www.edweek.org/media/2011-07-19artsintegration.pdf>
- Shiffman, C. D., Riggan, M., Massell, D., Goldwasser, M., & Anderson, J. (2008). Channeling adaptation: The role of design in enactment patterns. In J. A. Supovitz & E. H. Weinbaum (Eds.), *The implementation gap: Understanding reform in high schools* (pp. 46–67). New York, NY: Teachers College Press.
- U.S. Department of Health and Human Services, Administration for Children and Families. (2010). *Head Start Impact Study. Final report*. Washington, DC: Author. Retrieved from <http://files.eric.ed.gov/fulltext/ED507845.pdf>
- Weinbaum, E., & Supovitz, J. (2010, April). Planning ahead: Making program implementation more predictable. *Phi Delta Kappan*, 91(7), 68–71.
- Wilson, S. (2013). Professional development for science teachers. *Science*, 340, 310–313.
- Yoon, K. S., Duncan, T., Lee, S. W.-Y., Scarloss, B., & Shapley, K. L. (2007). *Reviewing the evidence on how teacher professional development affects student achievement* (Issues & Answers Report, REL 2007–No. 033). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. Retrieved from [http://ies.ed.gov/ncee/edlabs/regions/southwest/pdf/rel\\_2007033\\_sum.pdf](http://ies.ed.gov/ncee/edlabs/regions/southwest/pdf/rel_2007033_sum.pdf)

## Appendix A. Examples of Arts Integration Routines Observed in Classroom Observations

### Example 1: Understanding Patterning Using Movement in Space

1. The teaching artist introduces a movement concept—kinesphere or space bubble. This will be the linking concept to the math students will study today.
2. Students are led in a movement activity in which they are asked to demonstrate how to grow or shrink their space bubbles in place and by moving as a class in an oval around the room.
3. Students review vocabulary that has been introduced and discuss how a change in size affects their bodies as they grow the bubble larger and shrink it smaller. Music is played using soft and loud levels.
4. The teaching artist reviews the vocabulary word (kinesphere) with students, and the class claps the number of syllables.
5. The teaching artist shows laminated cards with pictures showing bubbles of different sizes. Students discuss what it means to combine pictures in a pattern, which results in a display of bubbles of different sizes: small, big, small, big.
6. Students discuss what is meant by a pattern (ABAB), and they act out the patterns when the components are numbers (1 is small, 10 is large).
7. The teacher leads the class in activities about patterns using small and large letters and colors. Students create patterns and are asked to say the patterns using math symbols, letters, or colors.

### Example 2: Coffee Can Theater-Dramatization of Story/Understanding of Sets

1. Students have learned the story *Rooster's Off to See the World*. In this story, groups of animals (1 rooster, 2 cats, 3 frogs, 4 turtles, and 5 fish) go on a journey. When it gets late and dark, they turn around and return to their homes.
2. The teaching artist reminds the students of the story and the animals, reviewing the type of animal, their sounds, and why groups of animals belong together.
3. Artist and teacher reinforce the math concept by singing a song: let's sort, let's sort, who is the same? Teacher, artist, and students count the number of animals in each set.
4. Teacher and artist pass out sounds written on cards and give them out to students. Students are asked to sort themselves by the card they have and then make the sounds.
5. Then students are asked: Which group has more? If there is a difference of opinion, what do we do? Count! If students have difficulty identifying the difference when they only see the numbers on cards, they are asked to draw circles on a board/card, count the animals, and show how you can see the difference.
6. The artist acts out the Rooster story. The word *habitat* is introduced and used. Students are asked to tell what the habitat is for the frogs, the rooster, etc. Students count the number of animals in the different locations (the sets).

## Appendix B. Statistical Models for Impact Analyses

### Analysis Model for Assessing Program Impact on Opportunities for Arts Integration

Using classroom observation data, we assessed both the impact of the Wolf Trap professional development program on the likelihood that mathematics lessons provided opportunities for arts integration, and the impact of the program on the extent of arts integration implemented in the lessons observed. The impact on opportunities for arts integration at each wave of observation was assessed using the following logistic regression:<sup>11</sup>

$$\log\left(\frac{p_j}{1-p_j}\right) = \beta_0 + \beta_1(TREATMENT)_j + \beta_2(GROUP1)_j + \beta_3W_j + r_j$$

Where

- $p_j$  is the probability that the observed lesson taught by teacher  $j$  provided opportunities for arts integration,
- $(TREATMENT)_j$  is an indicator for treatment status of teacher  $j$ : 1 = treatment; 0 = control;
- $(GROUP1)_j$  is an indicator for the school group teacher  $j$  belonged to; 1 = Group 1; 0 = Group 2;
- $W_j$  is a vector of teacher background characteristics, including their years of teaching experience, primary teaching assignment (PK versus K), and race (white versus non-white); and
- $r_j$  is a random error associated with teacher  $j$ .

Coefficient  $\beta_1$  from the above model captures the impact of the professional development program on the likelihood (in logit) that a teacher's lesson provided opportunities for arts integration.

### Analysis Model for Assessing Program Impact on the Extent of Arts Integration

The analyses assessing the impact of the professional development program on the extent of arts integration were based on a two-level hierarchical linear model (HLM) that takes into account the clustering of teachers within schools. The analyses were performed separately for the overall measure of arts integration and each of the four individual measures of arts integration (i.e., Linking Arts With Math, Student Group Work in Art, Balanced Focus, and Artistic Product) using the following model:

Level 1 (teachers):

$$Y_{ij} = \beta_{0j} + \beta_{1j}W_{ij} + r_{ij}$$

---

<sup>11</sup> Ideally, this analysis should be conducted using a multilevel logit model that takes into account the clustering of teachers within schools. However, our analysis sample size is too small for the multilevel logit model, which had difficulty converging for some of the measures.



Where

- $Y_{ij}$  is the arts integration score for teacher  $i$  at school  $j$ ;
- $\beta_{0j}$  is the average arts integration score for teachers in school  $j$ , adjusted for teacher characteristics;
- $\beta_{1j}$  is the relationship between a given teacher characteristic and arts integration score at school  $j$ ; and
- $r_{ij}$  is a random error associated with teacher  $i$  at school  $j$ .

Level 2 (schools):

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(TREATMENT)_j + \gamma_{02}(GROUP1)_j + u_{0j}$$
$$\beta_{1j} = \gamma_{10}$$

Where

- $\gamma_{00}$  is the average arts integration score for Group 2 control schools;
- $\gamma_{01}$  is the difference in average arts integration score between treatment and control schools;
- $\gamma_{02}$  is the difference in average arts integration score between Group 1 and Group 2 schools;
- $\gamma_{10}$  is the average relationship between a given teacher characteristic and arts integration score; and
- $u_{0j}$  is a random error associated with school  $j$ .

Coefficient  $\gamma_{01}$  from the Level 2 model above captures the impact of the professional development program on the extent to which teachers used arts integration in their mathematics classrooms.

### **Analysis Models for Assessing Program Impact on Student Achievement**

Using EMDA data pooled across both Group 1 and Group 2 schools, we assessed both the one-year impact and the two-year impact of the professional development program on students' mathematics knowledge using a model specified similarly to the model for assessing the program's impact on the extent of arts integration. The HLM model for assessing the one-year impact is specified as follows:

Level 1 (students):

$$Y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + r_{ij}$$

Where

- $Y_{ij}$  is the EMDA score at the end of the first year of implementation (spring 2012 for Group 1 schools and spring 2013 for Group 2 schools) for student  $i$  at school  $j$ ;

- $X_{ij}$  is a vector of student characteristics, including their baseline EMDA scores, grade level at the beginning of the intervention (i.e., PK or K), gender, and race (white versus non-white), grand-mean centered;
- $\beta_{0j}$  is the average EMDA score for students in school  $j$ , adjusted for student characteristics;
- $\beta_{1j}$  is the relationship between a given student characteristic and the EMDA score at school  $j$ ; and
- $r_{ij}$  is a random error associated with student  $i$  at school  $j$ .

Level 2 (schools):

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(TREATMENT)_j + \gamma_{02}(GROUP1)_j + u_{0j}$$

$$\beta_{1j} = \gamma_{10}$$

Where

- $\gamma_{00}$  is the average EMDA score for Group 2 control schools;
- $\gamma_{01}$  is the difference in average EMDA score between treatment and control schools;
- $\gamma_{02}$  is the difference in average EMDA score between Group 1 and Group 2 schools;
- $\gamma_{10}$  is the average relationship between a given student characteristic and EMDA score; and
- $u_{0j}$  is a random error associated with school  $j$ .

The estimate of primary interest from the above model is  $\gamma_{01}$ , which captures the one-year impact of the Wolf Trap professional development program on students' mathematics achievement across both groups of schools. The two-year impact of the program was assessed similarly. Given the small number of schools in each school group (six schools in Group 1 and 12 schools in Group 2), we did not conduct the impact analyses for each school group separately.

## Appendix C. Interpretation of the Size of the Effects of Wolf Trap’s Professional Development Program

American Institutes for Research’s (AIR’s) recent evaluation of the professional development program developed by Wolf Trap as part of the Arts in Education—Model Development and Dissemination (AEMDD) grant found that the program had significant positive effects on students’ mathematics knowledge, with an effect size (standardized mean difference) of 0.17 for the first year of implementation and 0.21 for the second year of implementation.<sup>12</sup> The purpose of this appendix is to provide two types of information that may help interpret the size of these effects in substantively meaningful ways.

### Effects In Terms of Months of Learning

One way to interpret the effect sizes found in AIR’s evaluation is to compare them with the achievement growth that is expected to occur during a school year for an average student in the target grade. Ideally, such comparisons should be based on the expected annual achievement gains for the specific assessment used in the AIR study (i.e., Early Math Diagnostic Assessment [EMDA]). In the absence of such information for EMDA, information about annual achievement gains for other nationally normed standardized tests provides a reasonable alternative benchmark for converting the effects of the Wolf Trap professional development program into months of learning.

Specifically, based on national norming studies of three standardized achievement tests, Bloom, Hill, Black, and Lipsey (2008) found that the expected achievement growth in mathematics from kindergarten to Grade 1 is associated with an effect size of 1.14.<sup>13</sup> Assuming a nine-month school year and similar achievement growth for students in prekindergarten through Grade 1, an effect size of 0.17 (i.e., the first-year effect found in the Wolf Trap evaluation) would be equivalent to 1.3 months of learning, and an effect size of 0.21 (second-year effect) would be equivalent to 1.7 months of learning for students in the study.

### Effects of Other Early Childhood Interventions

Another way to understand the practical significance of the effects of the Wolf Trap professional development program is to compare them with the effects of other early childhood interventions, which may provide useful information for contextualizing the findings about the Wolf Trap program. Perhaps the most well-known impact study in the field of early childhood education is the Head Start Impact Study (U.S. Department of Health and Human Services, Administration for Children and Families, 2010), which assessed the effects of offering Head Start to 3- and 4-year-olds on a wide range of child outcomes based on a randomized controlled trial. Relying on data on about 4,700 children who applied for enrollment at a nationally representative sample of 380 Head Start centers, the study found that Head Start did not have a significant effect on any of the five measures of mathematics skills for the 4-year-old cohort in any of the three years

---

<sup>12</sup> Detailed information about the design and findings of the AIR evaluation is provided in the evaluation report (Ludwig & Song, 2015), which can be found online at

[http://www.wolftrap.org/~media/files/pdf/education/stem\\_fullstudy\\_aemdd\\_report\\_2015.pdf?la=en](http://www.wolftrap.org/~media/files/pdf/education/stem_fullstudy_aemdd_report_2015.pdf?la=en)

<sup>13</sup> Similar information is not available for prekindergarten to kindergarten.

examined. The study found that Head Start had a significant effect on one mathematics achievement measure in one of the four years examined for the 3-year-old cohort, with an effect size of 0.15. Tables C1 and C2 provide the effect sizes for all findings about Head Start’s impact on children’s mathematics achievement from the Head Start Impact Study, which are all smaller than the effects of the Wolf Trap program.

**Table C1. Effect Sizes of Head Start’s Effects on the Mathematics Skills of the 3-Year-Old Cohort, By Year**

Measures	Age 3 (Head Start Year)	Age 4	Kindergarten	First Grade
One-to-One Counting	0.06	0.02		
Applied Problems (WJIII)	0.15	0.03	-0.04	0.08
Quantitative Concepts (WJIII)			-0.05	0.05
Math Reasoning (WJIII)			-0.05	0.07
Calculation (WJIII)				0.00

Note: WJIII=Woodcock-Johnson III

**Table C2. Effect Sizes of Head Start’s Effects on the Mathematics Skills of the 4-Year-Old Cohort, By Year**

Measures	Age 4 (Head Start Year)	Kindergarten	First Grade
One-to-One Counting	0.08		
Applied Problems (WJIII)	0.12	0.01	0.04
Quantitative Concepts (WJIII)		-0.01	0.02
Math Reasoning (WJIII)		0.00	0.03
Calculation (WJIII)			0.07

Note: WJIII=Woodcock-Johnson III

The Head Start Impact Study was rated by the What Works Clearinghouse (WWC) as *meeting WWC evidence standards without reservations*; and a detailed intervention report can be found at [http://ies.ed.gov/ncee/wwc/pdf/intervention\\_reports/wwc\\_headstart\\_072815.pdf](http://ies.ed.gov/ncee/wwc/pdf/intervention_reports/wwc_headstart_072815.pdf). In addition to Head Start, we also reviewed research evidence released by the WWC for the effectiveness of other early childhood interventions. Of the more than 80 early childhood interventions that WWC has identified, five interventions other than Head Start had at least one study that met WWC evidence standards, either with or without reservations, and that examined mathematics achievement outcomes (see Table C3). Of those five interventions, two had much larger effects, two had similar effects, and one had a much smaller effect, relative to the effects of the Wolf Trap program. Detailed information about these five interventions and study findings can be found in the intervention reports released by the WWC at the links provided in Table C3.

**Table C3. Findings About the Effects of Selected Early Childhood Interventions on Children’s Mathematics Achievement**

Early Childhood Interventions Reviewed by the WWC	Research Base	Effect on Math Achievement
<p><i>Pre-K Mathematics</i> is a supplemental curriculum designed to develop the informal mathematical knowledge and skills of preschool children with content organized into seven units.  <a href="http://ies.ed.gov/ncee/wwc/pdf/intervention_reports/wwc_prekmath_121713.pdf">http://ies.ed.gov/ncee/wwc/pdf/intervention_reports/wwc_prekmath_121713.pdf</a></p>	<p>One study met WWC standards and one study met WWC standards with reservations.</p>	<p>Significant positive effect: ES = 0.50 across two studies</p>
<p><i>The Creative Curriculum® for Preschool, Fourth Edition</i> is an early childhood curriculum that focuses on project-based investigations as a means for children to apply skills and addresses four areas of development: social/emotional, physical, cognitive, and language.  <a href="http://ies.ed.gov/ncee/wwc/pdf/intervention_reports/wwc_creativecurriculum_030513.pdf">http://ies.ed.gov/ncee/wwc/pdf/intervention_reports/wwc_creativecurriculum_030513.pdf</a></p>	<p>One study met WWC standards and one study met WWC standards with reservations.</p>	<p>Nonsignificant effect: ES = 0.04 across two studies</p>
<p><i>Tools of the Mind</i> is an early childhood curriculum for preschool and kindergarten children, based on the ideas of Russian psychologist Lev Vygotsky. The curriculum is designed to foster children’s executive function, which involves developing self-regulation, working memory, and cognitive flexibility. Many activities emphasize both executive functioning and academic skills.  <a href="http://ies.ed.gov/ncee/wwc/pdf/intervention_reports/wwc_toolsofmind_091608.pdf">http://ies.ed.gov/ncee/wwc/pdf/intervention_reports/wwc_toolsofmind_091608.pdf</a></p>	<p>One study met WWC standards.</p>	<p>Nonsignificant effect: ES = 0.17</p>
<p><i>SRA Real Math Building Blocks PreK</i> (also referred to as <i>Building Blocks for Math</i>) is a supplemental mathematics curriculum designed to develop preschool children’s early mathematical knowledge through various individual and small- and large-group activities.  <a href="http://ies.ed.gov/ncee/wwc/pdf/intervention_reports/WWC_Building_Blocks_072307.pdf">http://ies.ed.gov/ncee/wwc/pdf/intervention_reports/WWC_Building_Blocks_072307.pdf</a></p>	<p>Two studies met WWC standards.</p>	<p>Significant positive effect: ES = 1.07 across two studies</p>
<p><i>Direct Instruction</i> is a family of interventions focusing on teaching techniques that are fast-paced, teacher-directed, and explicit with opportunities for student response and teacher reinforcement or correction.  <a href="http://ies.ed.gov/ncee/wwc/pdf/intervention_reports/WWC_Direct_Instruction_052107.pdf">http://ies.ed.gov/ncee/wwc/pdf/intervention_reports/WWC_Direct_Instruction_052107.pdf</a></p>	<p>One study met WWC standards with reservations.</p>	<p>Nonsignificant effect: ES = 0.19</p>

One thing to note is that studies of Head Start and the other five early childhood interventions reviewed by the WWC differ from AIR’s evaluation of the Wolf Trap program, not only in the sample and setting, but also in the mathematics assessment used. Therefore, although findings about the effects of those other early childhood interventions provide a useful context for understanding the findings about the Wolf Trap program, those findings are not directly comparable.



## ABOUT AMERICAN INSTITUTES FOR RESEARCH

Established in 1946, with headquarters in Washington, D.C., American Institutes for Research (AIR) is an independent, nonpartisan, not-for-profit organization that conducts behavioral and social science research and delivers technical assistance both domestically and internationally. As one of the largest behavioral and social science research organizations in the world, AIR is committed to empowering communities and institutions with innovative solutions to the most critical challenges in education, health, workforce, and international development.

## LOCATIONS

### Domestic

Washington, D.C.  
Atlanta, GA  
Baltimore, MD  
Chapel Hill, NC  
Chicago, IL  
Columbus, OH  
Frederick, MD  
Honolulu, HI  
Indianapolis, IN  
Naperville, IL  
New York, NY  
Rockville, MD  
Sacramento, CA  
San Mateo, CA  
Waltham, MA

### International

Egypt  
Honduras  
Ivory Coast  
Kyrgyzstan  
Liberia  
Tajikistan  
Zambia







**AIR**<sup>®</sup>

AMERICAN INSTITUTES FOR RESEARCH<sup>®</sup>

1000 Thomas Jefferson Street NW  
Washington, DC 20007-3835  
202.403.5000 | TTY 877.334.3499

**[www.air.org](http://www.air.org)**

*Making Research Relevant*