Arts Achieve: Impacting Student Success in the Arts

Final Evaluation Findings

Tara M. Mastrorilli, M.Phil., Susanne Harnett, Ph.D., Jing Zhu, Ph.D.

Metis Associates

Author's Note: The authors would like to acknowledge all the dedication and support of the *Arts Achieve* partners and all of their staff in this work: the New York City Department of Education Office of Arts and Special Projects, Studio in a School, ArtsConnection, the Dance Education Laboratory at the 92nd Street Y; the Weill Music Institute at Carnegie Hall, the Cooper Hewitt National Design Museum, and all the participating schools. This research was supported by two United States Department of Education grants: an Arts in Education Development and Dissemination grant and an Investing in Innovation grant. Opinions reflect those of the authors and do not necessarily reflect those of the granting agency.

Abstract

Arts Achieve: Impacting Student Success in the Arts was a five-year project (from 2010-2015), which involved a partnership between the New York City Department of Education (NYCDOE) and five of the city's premier arts organizations. *Arts Achieve* was guided by three over-arching goals: 1) to improve student achievement in the arts; 2) to enhance arts teacher practice; and 3) to integrate technology into arts classrooms.

In order to achieve these goals, Arts Achieve provided intensive and targeted professional development to arts teachers over a three-year period. Starting in 2011-2012, arts teachers formed art discipline-based professional learning communities (PLCs) to work together using a process of inquiry and action research that focused on reviewing student data and examining impact on current instructional practice. Additionally, each arts teacher was paired with a facilitator from the arts organizations to support them over the course of the project. The specific professional development activities included: on-site consultancies, assessment retreats, inter-visitations, and an online community. Participating arts teachers also were provided with resources to support this work, such as units of study and technology bundles.

Metis Associates, an independent research and evaluation firm, conducted a rigorous cluster randomized control trial study of *Arts Achieve* in order to examine its impact on arts teachers and their students, whereby 77 schools were assigned to treatment or status-quo control conditions by arts discipline (dance, music, theater, visual arts) and school level (elementary, middle, high). To measure students' arts achievement, in the planning year of the project, Benchmark Arts Assessments were developed in each arts discipline and school level. Findings with respect to the project's three main goals are presented from data collected during Years 1-3 of the evaluation.

Introduction

Teaching and learning in the arts, like any subject, requires the use of assessment to allow teachers to reflect on students' progress and adjust their instruction to meet students' needs. Assessment also provides students with feedback, which allows them to reflect on their own learning. More specifically, current research highlights the importance of balanced assessment (formative and summative) in instruction to support student learning and increase student achievement (Black & Wiliam, 1998; Hattie & Timperley, 2007; Kluger & DeNisi, 1996). In general, however, the arts lack access to high quality assessments to inform arts teachers about their students' knowledge and skills in a particular art form. The *Arts Achieve: Impacting Student Success in the Arts* project was conceptualized to address this issue and, in doing so, improve teacher instruction and student learning in the arts.

In spring 2010, the *Arts Achieve* project was funded by two United States Department of Education (USDOE) grants: the New York City Department of Education (NYCDOE) Office of Arts and Special Projects (OASP) received an Arts Education Model Development and Dissemination (AEMDD) grant and Studio in a School (STUDIO) received an Investing in Innovation (i3) grant to implement the program. The goal of *Arts Achieve* was to improve the quality of arts teachers' instruction through intensive and targeted professional development on the use of balanced assessment, leading to increases in students' arts achievement. To evaluate the efficacy of the *Arts Achieve* project, the project partners hired Metis Associates, a national evaluation and consulting firm, to conduct an experimental study. At the end of the first project year (2010-2011), which was used as a planning year, 77 NYC public schools were assigned to treatment or status-quo control conditions using a stratified (art discipline and school level)

random assignment process. In this article, we report final outcome data for the *Arts Achieve* project on arts teachers' instructional practices as well as on their students' arts achievement.

Theoretic Framework

The logic model, which is displayed in Figure 1, illustrates the theoretical framework of the *Arts Achieve* project. The project was based on the theory that when arts teachers participate in professional development, including being members of professional learning communities (PLCs), that focuses on action research and the use of balanced assessment strategies, the quality of their arts instruction will improve. In turn, the project hypothesizes that enhanced arts instruction will lead to improved student achievement in the arts. The paragraphs below provide an overview of research on these components, which guided the work of *Arts Achieve.*

Arts Achieve drew on current research around professional development for educators, which emphasizes the need for PLCs to develop the collective capacity of staff to work together in order to improve teacher practice and student learning. Through PLCs, teachers work together in teams, learning from each other in environments that are both school-based and job embedded (DuFour, Eaker, DuFour, 2005). The development of PLCs in the arts community is particularly needed, given that arts teachers can be the only staff members in their schools teaching in their content area, and many have reported that they do not feel integral to the faculty (Burnaford, 2009).

Several research studies provide robust support regarding the impact of PLCs on teacher practices, school culture, and student achievement. For example, Louis and Marks (1998) conducted a multi-site study in 24 schools on the impact of PLCs. Through classroom observations and interviews with teachers, they documented the presence of authentic pedagogy (i.e., pedagogy that emphasizes higher order thinking, the construction of meaning through

conversation, and the development of depth of knowledge that has value beyond the classroom) and examined the connection between the quality of classroom pedagogy and the existence of core characteristics of a PLC. This study found that the presence of a professional community in a school contributes to higher levels of social support for achievement and higher levels of authentic pedagogy. Bolam, McMahon, Stoll, Thomas, & Wallace (2005) examined survey data from 393 schools and interview-based case study data from 16 school sites. Both the survey and case study data revealed that when teachers work in PLCs, there are fundamental shifts in the ways that they approach their work, increasing their collaboration, reducing their feelings of isolation, and improving their overall morale.

There is also strong evidence that when teachers work in PLCs, their students experience increased achievement. Eight recent studies (Berry, Johnson, & Montgomery, 2005; Bolam et al., 2005; Hollins, Mclntyre, DeBose, Hollins, & Towner, 2004; Louis & Marks, 1998; Phillips, 2003; Strahan, 2003; Supovitz, 2002; Supovitz & Christman, 2003) all found positive relationships between teachers' participation in PLCs and student academic achievement. In these studies, results of student achievement gains varied with the strength of the PLC in the school (Bolam et al., 2005; Louis & Marks, 1998) or with the specific focus of the efforts of teams or small communities of teachers (Supovitz, 2002; Supovitz & Christman, 2003).

Underlying all of the work of the PLCs is an emphasis on action research and use of appropriate balanced assessment strategies to review and discuss student work and teacher practice. Action research involves teachers engaging in inquiry and reflection on their current practice and student work. In contrast to one-day professional development sessions, professional development that involves action research is more participant-driven and incorporates inquiry and reflection that occurs over a period of time. Zeichner (2003) notes that

these components are in alignment with the standards and guiding principles for professional development as set by national organizations (e.g., the American Federation of Teachers, the National Councils of Teachers of English and Mathematics) and academic scholars (e.g., Darling-Hammond & McLaughlin, 1996).

Studies have found that teachers who engage in action research report that they have higher self-esteem and confidence levels (Dadds, 1995; Loucks-Horsely, Hewson, Love, & Stiles, 1998), develop self-analysis skills that are applied to their teaching (Day, 1984), and become more aware of how they impact their students (Allen, Shockley, & Baumann, 1995). In a review of studies on four action research professional development programs, Zeichner (2003) reported that engaging in action research helps teachers create a more student-centered environment, in which the teachers focus on listening to and observing their students to influence instruction. He explained further that teachers begin to see the point of view of their students and allow them to have more input in the classroom.

In order to inform teachers' action research, balanced assessment is a critical tool to gather evidence on student progress. Current research emphasizes the need for teachers to use assessment as a tool to help gather student performance data and assess how they should target their instruction to meet students' needs (Gewertz, 2010; Stiggins, 2010). Assessment also provides feedback to students in order for them to shape their understanding and improve their learning. Formative assessment helps students answer the question "How am I doing?" as they are learning new material and then summative assessment answers the question "How did I do?" at the end of learning a unit (Shute, 2008; Stiggins, 2005).

Recent meta-analyses have documented the effectiveness on the use of assessment practices on student achievement (Black & Wiliam, 1998; Hattie & Timperley, 2007; Kluger &

DeNisi, 1996). Black and Wiliam (1998) examined 250 studies from the research addressing a range of student factors and teacher instructional practices, including formative assessment strategies. They concluded that formative assessment has a more profound effect on learning than do other typical educational interventions, finding effect sizes between .4 and .7. Moreover, they concluded that assessment practices have a stronger effect on low achieving students than on high achieving ones, as they are instrumental in developing meta-cognitive skills and enhancing motivation. Hattie and Timperley's (2007) research described the results of 12 meta-analyses that included 196 studies and 6,972 effect sizes. Notably, they found that the average effect size for the use of formative feedback was .79.

Arts Achieve Project Implementation

Arts Achieve included a partnership between the NYCDOE OASP, five of the city's premier arts organizations, and Metis Associates, the project's evaluator. The arts organizations and their particular arts focus for the project were as follows: STUDIO (lead partner, visual arts); ArtsConnection (theater); the Dance Education Laboratory at the 92nd Street Y (dance); the Weill Music Institute at Carnegie Hall (music), and the Cooper Hewitt National Design Museum (technology). Figure 2 provides a description of each of the partner organizations. Prior to the first year of project implementation, the NYCDOE and the partner organizations developed and piloted 12 Benchmark Arts Assessments, one in each arts discipline (dance, music, theater, and visual arts) and school level (elementary, middle, and high). The Benchmark Arts Assessments measure students' arts content knowledge and performance skills based on local arts standards

set in the NYCDOE *Blueprints for Teaching and Learning in the Arts*.[1] The development of the Benchmark Arts Assessments is described further in the Methods section.

The *Arts Achieve* project provided professional development to arts teachers over a three-year period to help them engage in action research and to learn to use the data from the Benchmark Arts Assessments and ongoing formative assessments. Starting in the 2011-2012 school year, the arts teachers formed art discipline-based PLCs to work together toward the goal of improving their practice and student learning. Teaching artists from the arts organizations were partnered with the participating arts teachers to work together over the course of the project. In contrast to an artist-in-residence program, the teaching artist's role was that of a facilitator, who helped support the arts teacher with project work, as well as plan for instruction and provide assistance in the classroom. The *Arts Achieve* in-service professional development offered during the implementation years included:

- *On-site Consultancies*: Each facilitator visited his/her partner school about twice a month over the course of the school year, for a total of 22 visits over the course of the school year. During each of these visits, the facilitator and arts teacher worked directly with two of the teacher's targeted art classes[2] and participated in a separate joint instructional planning meeting. While the specific activities during each visit were unique to the needs of the arts teacher, the facilitator's time in the classroom

[1] The *Blueprints* set clear standards for what students should know, understand, and be able to do in each of the four arts disciplines (dance, music, theater, and visual arts) as they move through the school system from Pre-K through 12th grade. They are based on National arts standards and support the NYS Standards for Arts Instruction. Scope and sequence of learning are identified on the *Blueprints* through five strands: Art Making, Literacy in the Arts, Making Connections, Community and Cultural Resources, and Careers and Lifelong Learning. Benchmarks are delineated at four levels – Grades two, five, eight, and 12.

[2] In elementary schools, fifth-grade classes are targeted. In middle schools, eighth-grade classes are targeted, though some sixth- and seventh-grade classes may be included. In high school, the targeted classes can include any grade from nine through 12.

could include observing, modeling, or co-teaching. During the planning time, the arts teacher and facilitator often discussed the teacher's action research and frequently used the time to review student work, examine formative and summative assessment data, and discuss the instructional implications of data.

- *Assessment retreats*: The facilitators and arts teachers met three times over the course of the school year as a full group to focus on the appropriate use of formative and summative assessment strategies and the use of data from the Benchmark Arts Assessments to inform instruction. The assessment retreats were facilitated by Dr. Heidi Andrade, an expert in formative assessment practices. These meetings also provided an opportunity for participants to meet within their discipline and grade level specific PLCs.
- *Inter-visitations*: Two school inter-visitations per year provided additional opportunities for participants to meet within their PLCs. During these inter-visitations, the arts teachers deepened the work of the PLCs by discussing student work and sharing best practices around instruction and use of formative assessment strategies. The initial visit was to a model school, while all subsequent visits were to other treatment schools.
- *Online Community*: To support ongoing collaboration and share ideas and resources, a social-networking site was created for the project using Ning. This online community allowed the *Arts Achieve* participants to share student work by posting pictures or videos, to upload units of study or links to websites, to network with other teachers in the project, to make announcements, and to archive project documents.

In addition, the *Arts Achieve* project provided participating arts teachers with resources to support their work. Resources included *Blueprints*-aligned exemplary units of study, links to websites and other sources, their students' results from the Benchmark Arts Assessments, and technology bundles. The technology bundles included three iPads and a projector to use in their classrooms. The arts teachers participated in a full-day of training and ongoing on how to operate and incorporate the iPads into their instruction. Through the incorporation of these additional resources, teachers had increased access to arts content, formative and summative assessment strategies, and instructional strategies designed to reach students through multiple venues.

The *Arts Achieve* project provided the participating facilitators with additional training during each implementation year of the project. At least three full days of formal workshops were held over the course of each year. During these trainings, the facilitators were provided with techniques and strategies on how to work with their partner arts teachers, how to apply formative assessment in the classroom, and how to use student data from formative assessments and the Benchmark Arts Assessments to improve instructional practice.

Purpose of Study

The literature on the positive impact of formative assessment, action research, and PLCs on teacher instruction and student achievement, as described above, provides ample evidence to support the *Arts Achieve* framework. However, the literature is limited in that it does not provide evidence on impacts on arts teacher instruction or student achievement in the arts, specifically. The purpose of this study was to extend and deepen this previous research, and to examine whether, and to what extent, providing teachers with high quality and intensive professional development can positively impact arts teachers and their students.

Overall, the *Arts Achieve* study addressed the following three research questions:

1. What is the nature of *Arts Achieve* implementation? What are the successes and challenges of *Arts Achieve* implementation?
2. What is the impact of the *Arts Achieve* project on arts teachers' *Blueprints* knowledge and instructional practices, including their use of the *Blueprints* standards and their interpretation and use of formative and summative assessment data?
3. What is the impact of the *Arts Achieve* project on students' arts achievement? Are there differential impacts of the *Arts Achieve* project by arts discipline and school level?

Methods

To evaluate the impact of the *Arts Achieve* project, Metis Associates designed a cluster randomized control trial, whereby schools were assigned to either the treatment or status-quo control condition. This design allowed for comparisons between teachers and students who participated in the treatment with those who did not. This section describes how the sample was selected, the instruments used to collect teacher and student background characteristics and outcomes, and the analyses used to address the study's research questions.

Sample

The *Arts Achieve* sample was selected using a two-stage process. First, NYCDOE public schools that met the basic eligibility requirements were recruited and then randomly selected to participate. To identify eligible schools, information about their arts programs was gathered through the 2009-2010 NYCDOE *Annual Arts Education Survey*.[3] To be eligible to participate, schools had to be part of the NYCDOE public school system and have an existing arts program in at least one arts discipline. Additionally, eligible *elementary* schools had to: 1) serve *all* grades 3-5; 2) offer at least 30 hours of instruction in one of the arts disciplines to all grades 3-5; and 3)

[3] Each spring all, NYCDOE schools are asked to complete the *Annual Arts Education Survey*. The survey collects information on schools' arts programs. More specifically, the survey includes questions about student participation in arts courses by arts discipline, arts activities conducted, and certified school-based arts teachers.

have a certified or a cluster teacher assigned to teach the arts. Eligible *middle* schools had to: 1) serve *both* grades 7 and 8; 2) offer at least one full year of instruction in an arts discipline to eighth-grade students; and 3) have an arts teacher on staff certified to teach the arts. Eligible *high* schools had to: 1) serve *any* grade 9-12; 2) offer at least one full year of instruction in an arts discipline to students in any grade 9-12; and 3) have an arts teacher on staff certified to teach the arts.

Two informational meetings in winter 2011 were held to describe the program and to recruit volunteer schools to participate in the study. Schools recruited for participation were blocked on two characteristics – arts discipline (dance, music, theater, and visual arts) and school level (elementary, middle, and high schools), thereby creating 12 blocks. In cases where there were more eligible schools than needed, schools were randomly selected to participate.

In the second stage of the selection process, schools were randomly assigned to the treatment or control group. In the first year, within each arts discipline block, eight schools were selected at the elementary level, six at the middle school level, and six at the high school level. At the elementary school level, five schools were assigned to the treatment group and three schools were assigned to the control group. At the middle and high school level, three schools were assigned to the treatment group and the three schools were assigned to the control group. In each implementation year, some schools dropped from the program, and in Years 2 and 3, schools were added. As schools were added, an even number of schools were selected for participation. Half of the schools were then randomly assigned to the treatment and half to the control group, according to the appropriate discipline and school level block. Tables 1-3 display the number of participating schools in the treatment and control groups at each discipline and school level block, including those that dropped out in each Year 1-3 and were added in Years 2

and 3. In Year 1, a total of 77 schools, 43 treatment and 34 control, participated in the full first year of implementation. In Year 2, a total of 83 schools (47 treatment and 36 control) completed participation; and in Year 3, a total of 78 schools (44 treatment and 34 control) completed participation. A total of 100 unique schools were included in the study. In total, 64 schools (37 treatment and 27 control) participated across all three years, and 36 (16 treatment and 20 control) participated for two or fewer years.

The treatment was mainly targeted towards the school's arts teacher in the arts discipline upon which the school was blocked. Within the treatment schools, the arts teachers' targeted classes in the treatment and control schools were selected based on grade (grade 5 in elementary, grade 8 in middle school,[4] and grades 9-12 in high school); year-long instruction (all target classes had to participate in year-long arts instruction); and (if multiple classes met these eligible requirements) scheduling convenience. The classes selected for assessment in the control schools were selected using the same criteria. Arts teachers and in the treatment schools participated in all the *Arts Achieve* professional development, whereas the arts teachers in the control schools only participated in evaluation activities.

In general, one arts teacher per school participated in the treatment, although there are some schools with more than one participating arts teacher. A total of 121 arts teachers participated in the project, including 65 arts teachers from the treatment schools and 56 arts teachers from the control schools. Thirty-one of the treatment arts teachers (48%) and 20 of the control teachers (36%) participated in the program for all three years.

More than 4,000 students participated in the program during each year of implementation. This included 4,066 students in Year 1 (2,046 treatment and 2,020 control),

[4] In some cases, if eight grade could not be targeted the treatment was geared toward grade six or seven.

5,402 students in Year 2 (2,638 treatment and 2,764 control), and 5,668 in Year 3 (2,812 treatment and 2,856 control. Tables 4-6 display the demographic profiles of participating students by group.

Thirty-three teaching artists, who all had previous experience working in schools, participated in the project as facilitators (nine dance, five music, nine theater, and ten visual arts). The majority of facilitators were partnered with only one arts teacher; however, some worked with two or three teachers.

Instruments

Ongoing data collection for *Arts Achieve* consisted of program documentation, surveys and focus groups with arts teachers, and the Benchmark Arts Assessments. Additionally, secondary data were collected from the NYCDOE; including students' background characteristics, and English Language arts achievement. This section describes the *Arts Achieve* data collection procedures and instruments.

Program documentation. To measure *Arts Achieve* implementation, program documentation is collected, including professional development materials, participant attendance, and observations of the trainings. Data on the number of on-site consultancies were also collected.

Arts teacher surveys. To collect information on arts teachers' knowledge and instructional skills, Metis and program staff developed surveys and administered them to the participating arts teachers in the treatment and control schools at the beginning (pre) and end (post) of the school year. The surveys asked questions about arts teachers' characteristics, including their years of teaching experience and certification in the designated arts discipline. The surveys also used Likert-scale questions to ask about arts teachers' knowledge and use of the

NYCDOE *Blueprints,* and use of formative and summative assessment data. Open-ended questions were also included to learn about the successes and challenges of implementing the *Arts Achieve* work in the school. A total of 42 arts teachers (35%) responded to both the baseline and post (spring 2014) surveys, including 30 treatment teachers (46%) and 12 control teachers (21%).

Composite scores of teachers' *Blueprints* knowledge, *Blueprints* use, and formative assessment strategies were calculated using multiple items on the arts teacher pre- and post-surveys. The *Blueprints* knowledge scale consisted of five items and had a Cronbach's alpha value of 0.77 on the post-survey. The use of *Blueprints* as well as the use of formative assessment strategies post-survey scales were both found to have high internal consistencies (4 items α=0.79, 3 items α=0.73, respectively).

Benchmark Arts Assessments. As described above, the Benchmark Arts Assessments were developed in the planning year and were used to measure the impact of *Arts Achieve* on students' arts achievement. Prior to the first year of implementation, teams were assembled to spearhead the creation of the assessments. The teams were led by the NYCDOE Arts Directors in the four arts forms, and also were comprised of school-based arts teachers, staff from the participating arts organizations, and NYCDOE experts in test and measurement. A total of 12 assessments were developed (three [one each for 5th grade, middle school, and high school] in each of the four arts disciplines [dance, music, theater, and visual arts]). The assessments were designed to measure the extent to which students have developed the knowledge and skills that are expected in a particular art form by fifth grade, middle school, and high school, according to the NYCDOE *Blueprints*. They also have been aligned with the Common Core Capacities in English Language Arts.

The Benchmark Arts Assessments each have multiple components, including performance and written sections, and include activities that address content knowledge, transferable concepts, and skills in the designated arts discipline. The assessments allow students to demonstrate their depth of knowledge in the content area through the analysis of performances, evaluation of other masterworks, and the creation of their own works of art. Question types include short-answer, fill-in-the-blank, multiple choice, and performance tasks. The majority of the questions were scored using a four-point scaled rubric.

Prior to each administration of the assessments, *Arts Achieve* staff provided training for adjudicators in administering and scoring the assessments. The adjudicators were current or retired NYCDOE arts teachers or facilitators. Separate trainings were conducted for each of the arts disciplines, but each included background on the context and purpose of the Benchmark Arts Assessments and training on the rubrics to score student work. Two adjudicators each administered and scored the assessments.

In spring 2011, the assessments were piloted in a sample of NYCDOE public schools that met the same requirements as the study schools, and the psychometric properties of the assessments were analyzed. Based on the results of the pilot, the assessments were found to have acceptable levels of reliability and validity. The assessments are administered to students in participating classes in the treatment and control schools in both the beginning (pre) and end (post) of the school year. Total scores for each of the assessments use a scale of 0 to 100.

Slight variations were made to the assessments in each of the implementation years. Tables 7-12 display the internal consistencies of each of the fall and spring Benchmark Arts Assessments. The Cronbach's alpha values all fall within the acceptable range, at .70 or above, with the exception of the middle school theater *musical theater design* task, which was .67.

Inter-rater reliabilities were calculated for each task scored using a rubric on each of the assessments. Across all three years of post-test items, 78% of the items fell at .31 or above and 23% fell at .70 or above. See Tables 13-16 for a summary of the interrater reliability data.

Secondary data. Data on treatment and control students' characteristics were obtained from the NYCDOE to use as covariates in the analysis models. Student characteristics include student demographics (including gender, race/ethnicity, English language learner status, special education status, and free/reduced priced lunch status), and average daily attendance. Students' scores on the New York State English Language Arts (NYS ELA) exam also were collected as a measure of student academic achievement. In elementary school and middle school, the NYS ELA assessment is administered to all students in grades three through eight in the spring of each school year.

Data Analysis

Descriptive statistics were calculated, including frequencies, means, and standard deviations, on the attendance, survey, and arts achievement data. Three-level hierarchical linear modeling (HLM) was conducted to measure the impact of the *Arts Achieve* project on students' arts achievement as well as on teachers' instructional practice. Additionally, multiple regression analyses were conducted to measure the impact of the *Arts Achieve* project on arts teachers' use of technology. Potential confounding factors that have a relationship with the outcomes (e.g., student characteristics, ELA achievement) were included in both the HLM as well as the multiple regression analyses to reduce threats to the study's internal validity. Covariates included in the teacher outcome models included arts teachers' years of experience, arts discipline certification, and pre-survey composite scores. In the student arts achievement models, student demographics (i.e., gender, race, English language learner status, special education status, free

and reduced price lunch status, and average daily attendance) and prior achievement (i.e., spring 2011 NYS ELA exam score and pre-Benchmark Arts Assessment score) were used as covariates. All covariates included in the HLM multiple regression models were grand mean centered, except for the treatment indicator. For statistically significant impact effects, Glass's delta was calculated to obtain the impact's effect size.

Results

As stated above, the *Arts Achieve* project was based on the theory that participation in the *Arts Achieve* professional development would have positive impacts on arts teachers' knowledge and instructional skills and, in turn, on students' arts achievement. This section presents implementation and impact findings of the project following all three years of implementation,

Implementation Findings

As shown in the Arts Achieve logic model, activities included professional development for arts teachers intended to impact on their instruction in the classroom with students, as well as some school-wide activities, such as school arts teams meetings. For purpose of examining the fidelity of implementation (FOI) with which Arts Achieve was implemented, four components of implementation were defined. Component 1 consisted of attendance at the professional development workshops, including the facilitator trainings, annual kick-off meetings, assessment and instructional retreats, and inter-visitations. Component 2 of the program consisted of the completion of on-site consultancies. Component 3 was comprised of the classroom planning and instruction activities that occurred during the on-site consultancies, including the use of formative assessment and technology strategies in the classroom. Finally, Component 4 was comprised of school planning and programming activities, including facilitator, art teacher, and

school administrator attendance at the school arts team meetings and the proportion of participating students who received year-long instruction. The evaluation team worked with the project staff to define fidelity thresholds for each component at the participant or school level, as well as at the program level.

Component 1 was calculated based on eight indicators and ranged from 0 to 100 points. A fidelity threshold at the participant level was set at 78 points or more. Component 2 was calculated based on one indicator, which ranged from 0 to 100 points, and the fidelity threshold was a score of 80 or more. Component 3 was calculated based on nine indicators and ranged from 0 to 160 points, with a fidelity threshold of 138 points or more. Lastly, Component 4 was calculated based on four indicators, ranging from 0 to 40 points, with a threshold of 31 points or more. At least 75 percent of schools had to meet the fidelity threshold for each component in order for the program to meet fidelity.

Table 17 displays the program level fidelity results from all three years of implementation. In each year of implementation, the program implemented the on-site consultancies with fidelity. Although the majority of participating *Arts Achieve* treatment schools implemented the on-site consultancies with fidelity, there was some slight variation from year to year. In Year 1, 91 percent of the treatment schools implemented the on-site consultancies with fidelity, as compared to 89 percent in Year 2, and 96 percent in Year 3. In all three years, however, the program did not meet the fidelity thresholds for any of the other three components, including professional development, classroom planning and instruction, and school planning and arts programming. For the professional development component, 42 percent of treatment schools met the participant-level fidelity in Year 1, 28 percent in Year 2, and 34 percent in Year 3. About half (47%) of treatment schools implemented the classroom planning and instruction

component with fidelity in Year 1, 57 percent in Year 2, and 48 percent in Year 3. For the school planning and arts programming, 49 percent of treatment schools met fidelity in Year 1, as compared to 23 percent in Year 2, and 18 percent in Year 3.

The on-site consultancies were the component over which the project had most direct control. Therefore, it is perhaps not surprising that the level of fidelity was highest for this component. Moreover, the *Arts Achieve* implementation of the professional development, classroom planning and instruction, and school planning and arts programming components confronted challenges that are not atypical to programs implemented in high-need, urban schools. The participating art teachers and administrators experienced numerous demands on their time, and oftentimes, art can take a backseat to content areas that are subject to high-stakes testing. As a result, art teachers and administrators often found it difficult to leave the building for professional development workshops. Administrators also found it difficult to hold at least three school arts team meetings throughout the school year.

In terms of classroom planning and instruction, the art teachers often found it challenging to focus on planning when the facilitator was at the school site, due to other on-site demands placed on the arts teachers. Furthermore, while most of the facilitator and teacher partnerships were strong, a few suffered from trust issues, teacher reluctance to change their practice, and/or personality conflicts all of which could have affected implementation. Lastly, in Year 3, the final year of implementation, a fair amount of attention was paid toward documenting the *Arts Achieve* work for dissemination efforts. According to several of the arts teachers and facilitators, they felt it took away from the primary focus of implementation, which may explain the low fidelity scores in Year 3.

Despite challenges in achieving fidelity of implementation for the project, many participants were highly positive about the impact of the program. For example, arts teachers were very enthusiastic about having the opportunity to collaborate with other arts teachers working in the same discipline and grade level. As one arts teacher noted, "The best part of this grant is meeting with other art teachers to share ideas, concerns, thoughts, and triumphs." Other arts teachers reported that the project increased the rigor of their work, making their work with students better aligned with arts standards. As one teacher put it, "My principal goal this year was to create an arts program which focuses on rigorous curriculum and authentic assessments. Arts Achieve has opened the door for this work. I have a clearer understanding of what is needed to implement arts programs that are aligned to high student content and academic achievement standards." Along these lines, another arts teacher indicated, "The Arts Achieve project has greatly improved my assessment practices. I now use clearer rubrics, and I introduce them significantly earlier in the learning process." Most importantly, however, many of the teachers reported that they saw the impact that the project had on their students. As one dance teacher noted, "[Students] have come to expect that their work will be "recorded." It makes them feel empowered and that the work they are doing is important. They have progressed from looking at themselves only as individual dancers/choreographers to looking at themselves as part of an ensemble."

Impact Findings

Improving student achievement in the arts. In each year of implementation, students in the treatment schools made greater gains in their arts achievement than participating students in the control schools. The results of HLM analyses indicated that participating students in the treatment schools made significantly greater gains in their arts achievement than participating students in the control schools, after controlling for student, teacher, and school background

characteristics in Years 1, 2, and 3 of implementation. The gains between the students in the treatment and control schools were not statistically significant in the third year of implementation. The effect Year 1 effect size was 0.28, the Year 2 effect size was 0.20, and the Year 3 effect size was 0.09. See Table 18 for these results.

To determine if there were differences in the impact of *Arts Achieve* by arts discipline, separate HLM analyses were conducted for the appropriate subsample the each implementation year. At the end of the first year of implementation, after controlling for student, teacher, and school characteristics, the results indicate that students in the music and theater treatment schools scored higher on the Benchmark Arts Assessments than participating students in the music and theater control schools. The Year 1 music effect size was 0.28 and the Year 1 theater effect size 0.43. Similar results were evident in music in the second implementation year. After controlling for student, teacher, and school characteristics, there was a statistically significant effect of *Arts Achieve* in the music treatment schools, with an effect size of 0.47. In the third year of implementation, the gains between the students in the treatment and control schools were not statistically significant by discipline. See Table 19 for these results.

Enhancing arts teacher practice. *Arts Achieve* aimed to improve arts teacher practice by increasing their use of formative assessment strategies to identify gaps in students' knowledge and skills, aligning their content with the *Blueprints*, and improving their use of technology for instruction and assessment. To measure whether change occurred in teacher practice, arts teachers completed annual pre/post surveys, through which they self-reported the frequency with which they used formative assessment strategies (i.e., teacher descriptive feedback, peer feedback, student self-reflection, and student self-assessment), taught the strands of the *Blueprints* (e.g., developing arts-based literacy, making connections through the art form), and

used technology (e.g., iPads, Smart board) in instruction. Rating scales ranged from 1 (lowest use) to 6 (highest use). An average across items was calculated to create a composite measuring arts teachers' instructional practice.

Multiple regression analyses of the instructional practice data indicate that there was a statistically significant effect of *Arts Achieve* on arts teachers' instructional practice in both Years 1 and 2 of implementation. In Year 3 of implementation, although the arts teachers in the treatment schools made greater gains than the arts teachers in the control schools, these gains were not statistically significant. The analyses controlled for teacher background characteristics and also adjusted for lack of baseline equivalence between the treatment and control group. The effect Year 1 effect size was 0.53, the Year 2 effect size was 0.59, and the Year 3 effect size was 0.27. See Table 20 and Figure 3 for these results.

Data from interviews with arts teachers provided evidence to support the positive gains shown in the analyses. One teacher explained that the project allowed her to perfect her craft as a teacher. Another teacher explained that through his review of the data, he knows where his students' weaknesses are, and they are aligned with the instructional areas in which he did not place focus on over the course of the year. Yet another teacher summarized that because of *Arts Achieve* she now has a better understanding of formative assessment, which has allowed her to teach in such a way that her students have become more responsible for their own learning. Overall, teachers were most positive about the PLCs that they participated in through the project, as well as having the facilitators, who served as "critical friends" in their classrooms. They explained that the project addressed the commonly felt isolation that arts teachers feel and often catapulted them to being leaders in their buildings, as the work was aligned with the most current

educational discussions about performance assessment, using data to inform instruction, formative strategies, and assessing the Common Core.

A set of post-implementation observations of participating arts teachers was conducted in the fifth and final year of the project. Data from these observations indicated that many of the arts teachers were continuing to use formative assessment and technology practices that they learned through the grant. While most teachers were not able to maintain their connections with their PLC groups, several had spread the practices they learned to all classes they teach and were able to turnkey learnings to other teachers at their schools. Many also were involved in other arts initiatives and were continuing to improve their knowledge and practice.

Integrating technology into arts classrooms. *Arts Achieve* teachers received technology bundles to use in their arts classrooms. Teachers were expected to use the iPads in various ways, including: taking pictures of student work; taking video to record students' performances and reflections; and using applications for instruction. The use of technology over the course of the three years of implementation was tracked using the implementation logs, which the facilitator and arts teacher pairs completed together during each school visit. The pair indicated if technology (i.e., iPad, desktop/laptop, Smart Board, video camera, still camera, sound system, projector or other device) was used in a given lesson.

According to data gathered through the logs, technology was used during 60 percent of the lessons in the first year of implementation. The percentage of use increased to 74 percent of lessons in the second year and 76 percent in the third year. Data were also analyzed for use of the iPads specifically, which were distributed to the *Arts Achieve* teachers in the winter of the first implementation year. By the end of that year, iPads were used for instruction during 24 percent of lessons, with percentages increasing to 46 percent the second year and 57 percent in the third.

Discussion

Arts Achieve was an extremely ambitious project that served an important unmet need in arts education. The project was designed not only to enable the NYC DOE to take the next step following their development of the *Blueprints* standards—to assess whether students are meeting these standards—but also to provide professional development and intensive supports to arts teachers so they might identify the gaps in student learning and have the instructional tools to fill them. The project was highly successful in many ways, including the development of high-quality, valid and reliable Benchmark Arts Assessments that may be used by arts teachers throughout NYC and beyond; and exemplar formative assessment practices, unit and lesson plans, and professional development models. Furthermore, data from the rigorous experimental design also indicate that the project had significant impact on the instructional practices of participating teachers and the arts achievement of their students. While the first two years of *Arts Achieve* show extremely strong results, the data indicate that the effect in Year 3 was not as strong. There are many possible explanations for the lack of significant findings in the final year of implementation. For example, it may be due to the greater amount of time that teachers and facilitators spent on documenting the *Arts Achieve* work in the final year as compared to the first two years. Though these efforts were extremely useful for replication efforts, several arts teachers and facilitators reported that it took away from implementation efforts. Additionally, it may be that the lack of findings regarding teachers' instructional practice is due to low response rates. Survey data collection from the participating treatment and control teachers was often challenging, becoming more so over time.

References

Allen, J.B., Shockley, B.B., Baumann, J.F. (1995). Gathering 'round the kitchen table': teacher inquiry in the NRRC school research consortium. *Reading Teacher, 48*, 526-529.

Berry, B., Johnson, D., & Montgomery, D. (2005). The power of teacher leadership. *Educational Leadership, 62*, 56-60.

Black, P. & Wiliam, D. (1998). *Inside the Black Box: Raising Standards Through Classroom Assessment*. *Phi Delta Kappan, 80*, 139-148.

Bolam, R., McMahon, A., Stoll, L., Thomas, S., & Wallace, M. (2005). *Creating and sustaining effective professional learning communities*. University of Bristol, UK.

Burnaford, G. (2009). A study of professional development for arts teachers: Building curriculum, community, and leadership in elementary schools. *Journal for Learning through the Arts*, 5, 1-45.

Dadds, M. (1995). *Passionate Inquiry and School Development*. Falmer Press: London, UK.

Day, C. (1984). Teachers' thinking intentions and practice: An action research perspective. In R. Halkes & J. Olson (Eds.) *Teacher Thinking*. Swets & Zeitlinger: Lisse, NL.

DuFour, R., Eaker, R. & DuFour, R. (2005). *On common ground: The power of professional learning communities.* Bloomington, IN: Solution Tree

Gewertz, C. (2010). Experts Lay Out Vision for Future Assessments. *Education Week, 29 (23)*. Retrieved from http://www.edweek.org/ew/articles/2010/02/23/23assessment.h29.html?tkn=VY%5bFA3BBpVhGN9TxkWdYT%2Byv0d9k%2Fn0QJf9z&cmp=clpedweek.

Hattie, J. & Timperley, H. (2007). The Power of Feedback. *Review of Educational Research* 77.1, 81-112.

Hollins, E.R., Mclntyre, LR, DeBose, C., Hollins, K.S., & Towner, A. (2004). Promoting a self-sustaining learning community: Investigating an internal model for teacher development. *International Journal of Qualitative Studies in Education*, 17, 247-264.

Kluger & DeNisi (1996). The Effects of Feedback Interventions on Performance: A Historical Review, a Meta-Analysis, and a Preliminary Feedback Intervention Theory. *Journal of Applied Psychology*, 254-284.

Loucks-Horsely, S., Hewson, P., Love, P. & Stiles, K. (1998) *Designing Professional Development for Teachers of Science and Mathematics*. Corwin Press: Thousand Oaks, CA.

Louis, K.S. & Marks, H.M. (1998). Does professional learning community affect the classroom? Teachers' work and student experiences in restructuring schools. *American Journal of Education,* 106, 532-575.

Phillips, J. (2003). Powerful learning: Creating learning communities in urban school reform. *Journal of Curriculum and Supervision, 18,* 240-258.

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research, 78,* 153-189.

Stiggins, R. (2010). Essential formative assessment competencies for teachers and school leaders. *Handbook of Formative Assessment.* New York: Routledge

Strahan, D. (2003). Promoting a collaborative professional culture in three elementary schools that have beaten the odds. *The Elementary School Journal, 104*, 127-146.

Supovitz, J. A. (2002). Developing communities of instructional practice. *Teachers College Record, 104*, 1591-1626.

Supovitz, J. A., & Christman, J. B. (2003). Developing communities of instructional practice: Lessons from Cincinnati and Philadelphia. *CPRE Policy Briefs*, #RB-39.

Zeichner, K. (2003). Teacher research as professional development for p-12 educators in the U.S. *Educational Action Research*, 11, 301-325.

Table 1. Number of Participating Schools 2011-2012, including Drop Outs

	Treatment (*Arts Achieve* Participants)			**Control**		
	N Schools Started	N Schools Dropped	Total N Schools Completed	N Schools To Start	N Schools Dropped	Total N Schools Completed
Dance	11	0	11	9	0	9
Music	11	0	11	9	0	9
Theater	11	1	10	9	1	8
Visual Arts	11	0	11	8	0	8
Total	**44**	**1**	**43**	**35**	**1**	**34**

Table 2. Number of Participating Schools in 2012-2013, including Additions and Drop Outs

	Treatment (*Arts Achieve* Participants)				**Control**			
	N Schools Started	N Schools Dropped	N Schools Added	Total N Schools Completed	N Schools Started	N Schools Dropped	N Schools Added	Total N Schools Completed
Dance	11	2	3	12	9	2	3	10
Music	11		1	12	9	1	1	9
Theater	10	3	4	11	8	3	4	9
Visual Arts	11	0	1	12	8	1	1	8
Total	**43**	**5**	**9**	**47**	**34**	**7**	**9**	**36**

Table 3. Number of Participating Schools in 2013-2014, including Additions and Drop Outs

	Treatment (*Arts Achieve* Participants)				**Control**			
	N Schools Started	N Schools Dropped	N Schools Added	Total N Schools Completed	N Schools Started	N Schools Dropped	N Schools Added	Total N Schools Completed
Dance	12	3	2	11	10	2	2	10
Music	12	0	0	12	9	0	0	9
Theater	11	5	3	9	9	4	3	8
Visual Arts	12	0	0	12	8	1	0	7
Total	47	8	5	44	36	7	5	34

Table 4. Demographic Profile of Participating Students by Group in Year 1 of Implementation

Student Characteristic	Treatment	Control
N Students	2,046	2,020
Female	54.8%	56.5%
Race		
White	14.9%	14.6%
African American	28.6%	31.5%
Asian	18.5%	17.7%
Hispanic	37.0%	35.7%
Other	0.93%	0.45%
English Language Learner	14.0%	8.7%
Special Education	13.9%	11.4%
Free-Reduce Priced Lunch	83.8%	79.3%
Average Daily Attendance (SD)	93.5 (10.58)	92.8 (10.21)
Spring 2011 NYSELA Proficiency[a]	50.4%	55.6%

[a] Students' spring 2011 NYSELA proficiency level was not available for all participating students. There were 1,618 treatment students with spring 2011 NYSELA scores and 1,467 control students with scores.

Table 5. Demographic Profile of Participating Students by Group (Year 2 of Implementation)

Student Characteristic	Treatment	Control
N Students	2,638	2,764
Female	55.4%	53.9%
Race		
White	19.2%	13.6%
African American	28.2%	26.7%
Asian	17.3%	19.6%
Hispanic	34.6%	38.7%
Other	0.7%	1.4%
English Language Learner	12.6%	14.6%
Special Education	13.2%	12.0%
Free-Reduce Priced Lunch	82.6%	80.6%
Average Daily Attendance (SD)	93.4 (9.97)	93.6 (9.96)
Spring 2012 NYSELA Proficiency[a]	53.4%	57.1%

[a] Students' spring 2012 NYSELA proficiency level was not available for all participating students. There were 2,592 treatment students with spring 2012 NYSELA scores and 2,658 control students with scores.

Table 6. Demographic Profile of Participating Students by Group in Year 3 of Implementation

Student Characteristic	Treatment	Control
N Students	2,812	2,856
Female	56.4%	57.3%
Race		
White	17.2%	15.9%
African American	29.4%	28.3%
Asian	18.4%	19.0%
Hispanic	34.0%	35.6%
Other	1.0%	1.2%
English Language Learner	11.3%	10.9%
Special Education	13.1%	13.8%
Free-Reduce Priced Lunch	70.7%	68.8%
Average Daily Attendance (SD)	93.6 (8.6)	93.8 (8.8)
Spring 2013 NYSELA Proficiency[a]	32.2%	37.5%

[a] Students' spring 2013 NYSELA proficiency level was not available for all participating students. There were 2,573 treatment students with spring 2013 NYSELA scores and 2,635 control students with scores.

Table 7. Internal Consistencies of the Benchmark Arts Assessments in Fall 2011 and Spring 2012

Benchmark Arts Assessment			Fall 2011			Spring 2012		
Art Discipline	School Level	Subtest[a]	N Cases	N Items	α	N Cases	N Items	α
Dance								
	Elementary		215	16	0.87	286	16	0.85
	Middle		226	15	0.83	214	15	0.88
	High		149	19	0.84	120	19	0.84
Music								
	Elementary	Instrumental	51	15	0.74	88	15	0.76
		Vocal	398	15	0.76	203	15	0.72
	Middle		306	23	0.89	319	23	0.86
	High		277	26	0.94	225	26	0.83
Theater								
	Elementary	Playwriting	115	12	0.81	127	12	0.85
		Costume Design	154	13	0.73	85	13	0.83
	Middle	Acting - Actors	59	9	0.87	65	9	0.81
		Musical Theater - Actors	28	7	0.83	25	7	0.77
		Playwriting - Actors	26	7	0.89	18	7	0.78
		Acting - Design	6	9	–	1	7	–
		Musical Theater - Design	11	7	0.67	5	7	–
		Playwriting - Design	3	7	–	0	7	–
	High	Staging	92	10	0.90	132	10	0.89
		Design Choices	117	11	0.89	39	11	0.77
Visual Arts								
	Elementary		525	33	0.89	441	29	0.85
	Middle		310	28	0.90	228	25	0.88
	High		203	18	0.88	183	18	0.87

[a] The music elementary assessment and each of the theater assessments asked students to choose between tasks on the assessments. In these cases, the internal consistencies are presented for each of the subtests.

Table 8. Internal Consistencies of the Benchmark Arts Assessments in Fall 2012 and Spring 2013

Benchmark Arts Assessment			Fall 2012			Spring 2013		
Art Discipline	School Level	Subtest[a]	N Cases	N Items	α	N Cases	N Items	α
Dance								
	Elementary		276	16	0.87	388	16	0.83
	Middle		259	15	0.77	307	15	0.80
	High		223	19	0.83	233	19	0.81
Music								
	Elementary	Instrumental	118	19	0.77	86	19	0.85
		Vocal	56	19	0.75	193	19	0.82
	Middle		391	13	0.70	387	15	0.79
	High		204	22	0.85	289	22	0.83
Theater								
	Elementary	Playwriting	238	15	0.79	263	15	0.79
		Costume Design	182	18	0.83	131	18	0.87
	Middle	Acting - Design	121	24	0.84	41	24	0.82
		Playwriting - Design	51	22	0.82	24	22	0.88
		Acting - Staging	49	24	0.83	54	24	0.91
		Playwriting - Staging	22	22	0.85	41	22	0.91
	High	Actor's Choices	149	11	0.91	49	11	0.91
		Design Choices	15	11	0.87	83	11	0.89
Visual Arts								
	Elementary		435	30	0.85	440	30	0.87
	Middle		275	25	0.89	293	25	0.87
	High		301	18	0.85	318	18	0.90

[a] The music elementary assessment and each of the theater assessments asked students to choose between tasks on the assessments. In these cases, the internal consistencies are presented for each of the subtests.

Table 9. Internal Consistencies of the Benchmark Arts Assessments in Fall 2013 and Spring 2014

Benchmark Arts Assessment			Fall 2013			Spring 2014		
Art Discipline	School Level	Subtest[a]	N Cases	N Items	α	N Cases	N Items	α
Dance								
	Elementary		408	16	0.83	397	16	0.85
	Middle		221	17	0.90	320	17	0.90
	High		280	21	0.86	210	21	0.81
Music								
	Elementary		215	17	0.80	315	17	0.78
	Middle		289	25	0.85	288	25	0.86
	High		191	24	0.90	200	24	0.77
Theater								
	Elementary		278	24	0.93	372	24	0.91
	Middle	Acting - Design	34	24	0.73	97	24	0.89
		Playwriting - Design	56	24	0.71	67	24	0.89
		Acting - Staging	32	24	0.71	32	24	0.88
		Playwriting - Staging	25	24	0.70	27	24	0.93
	High	Staging	88	12	0.87	76	12	0.79
		Design Choices	65	12	0.83	33	12	0.80
Visual Arts								
	Elementary		426	30	0.85	431	32	0.87
	Middle		267	31	0.86	206	32	0.89
	High		265	15	0.85	267	16	0.85

[a] The theater assessments asked students to choose between tasks on the assessments. In these cases, the internal consistencies are presented for each of the subtests.

Table 10. Internal Consistencies of the Benchmark Arts Assessments in Fall 2011 and Spring 2012

Benchmark Arts Assessment			Fall 2011			Spring 2012		
Art Discipline	School Level	Subtest[a]	N Cases	N Items	α	N Cases	N Items	α
Dance								
	Elementary		215	16	0.87	286	16	0.85
	Middle		226	15	0.83	214	15	0.88
	High		149	19	0.84	120	19	0.84
Music								
	Elementary	Instrumental	51	15	0.74	88	15	0.76
		Vocal	398	15	0.76	203	15	0.72
	Middle		306	23	0.89	319	23	0.86
	High		277	26	0.94	225	26	0.83
Theater								
	Elementary	Playwriting	115	12	0.81	127	12	0.85
		Costume Design	154	13	0.73	85	13	0.83
	Middle	Acting - Actors	59	9	0.87	65	9	0.81
		Musical Theater - Actors	28	7	0.83	25	7	0.77
		Playwriting - Actors	26	7	0.89	18	7	0.78
		Acting - Design	6	9	–	1	7	–
		Musical Theater - Design	11	7	0.67	5	7	–
		Playwriting - Design	3	7	–	0	7	–
	High	Staging	92	10	0.90	132	10	0.89
		Design Choices	117	11	0.89	39	11	0.77
Visual Arts								
	Elementary		525	33	0.89	441	29	0.85
	Middle		310	28	0.90	228	25	0.88
	High		203	18	0.88	183	18	0.87

[a] The music elementary assessment and each of the theater assessments asked students to choose between tasks on the assessments. In these cases, the internal consistencies are presented for each of the subtests.

Table 11. Internal Consistencies of the Benchmark Arts Assessments in Fall 2012 and Spring 2013

Benchmark Arts Assessment			Fall 2012			Spring 2013		
Art Discipline	School Level	Subtest[a]	N Cases	N Items	α	N Cases	N Items	α
Dance								
	Elementary		276	16	0.87	388	16	0.83
	Middle		259	15	0.77	307	15	0.80
	High		223	19	0.83	233	19	0.81
Music								
	Elementary	Instrumental	118	19	0.77	86	19	0.85
		Vocal	56	19	0.75	193	19	0.82
	Middle		391	13	0.70	387	15	0.79
	High		204	22	0.85	289	22	0.83
Theater								
	Elementary	Playwriting	238	15	0.79	263	15	0.79
		Costume Design	182	18	0.83	131	18	0.87
	Middle	Acting - Design	121	24	0.84	41	24	0.82
		Playwriting - Design	51	22	0.82	24	22	0.88
		Acting - Staging	49	24	0.83	54	24	0.91
		Playwriting - Staging	22	22	0.85	41	22	0.91
	High	Actor's Choices	149	11	0.91	49	11	0.91
		Design Choices	15	11	0.87	83	11	0.89
Visual Arts								
	Elementary		435	30	0.85	440	30	0.87
	Middle		275	25	0.89	293	25	0.87
	High		301	18	0.85	318	18	0.90

[a] The music elementary assessment and each of the theater assessments asked students to choose between tasks on the assessments. In these cases, the internal consistencies are presented for each of the subtests.

Table 12. Internal Consistencies of the Benchmark Arts Assessments in Fall 2013 and Spring 2014

Benchmark Arts Assessment			Fall 2013			Spring 2014		
Art Discipline	School Level	Subtest[a]	N Cases	N Items	α	N Cases	N Items	α
Dance								
	Elementary		408	16	0.83	397	16	0.85
	Middle		221	17	0.90	320	17	0.90
	High		280	21	0.86	210	21	0.81
Music								
	Elementary		215	17	0.80	315	17	0.78
	Middle		289	25	0.85	288	25	0.86
	High		191	24	0.90	200	24	0.77
Theater								
	Elementary		278	24	0.93	372	24	0.91
	Middle	Acting - Design	34	24	0.73	97	24	0.89
		Playwriting - Design	56	24	0.71	67	24	0.89
		Acting - Staging	32	24	0.71	32	24	0.88
		Playwriting - Staging	25	24	0.70	27	24	0.93
	High	Staging	88	12	0.87	76	12	0.79
		Design Choices	65	12	0.83	33	12	0.80
Visual Arts								
	Elementary		426	30	0.85	431	32	0.87
	Middle		267	31	0.86	206	32	0.89
	High		265	15	0.85	267	16	0.85

[a] The theater assessments asked students to choose between tasks on the assessments. In these cases, the internal consistencies are presented for each of the subtests.

Table 13. Counts and Percentages of Inter-rater Reliability for All Years (2011-2014)

Kappa ≤.3	Kappa = .31 to .69	Kappa = .7 or Above	Total Valid Kappa
132	339	142	613
22%	55%	23%	

Table 14. Counts and Percentages of Inter-rate Reliability for 2011-2012

Kappa ≤.3	Kappa = .31 to .69	Kappa = .7 or Above	Total Valid Kappa
11	94	100	205
5%	46%	49%	

Table 15. Counts and Percentages of Inter-rate Reliability for 2012-2013

Kappa ≤. .3	Kappa = .31 to .69	Kappa = .7 or Above	Total Valid Kappa
62	116	25	203
31%	57%	12%	

Table 16. Counts and Percentages of Inter-rate Reliability for 2013-2014

Kappa ≤. .3	Kappa = .31 to .69	Kappa = .7 or Above	Total Valid Kappa
59	129	17	205
29%	63%	8%	

Note: Counts reflect data collected during the spring semester only.

Table 17

Arts Achieve Fidelity of Implementation Results

Key Components	Definitions		Findings					
			Year 1		Year 2		Year 3	
	Measurement	Definition of fidelity at program level	% of schools meeting fidelity	Was project fidelity reached?	% of schools meeting fidelity	Was project fidelity reached?	% of schools meeting fidelity	Was project fidelity reached?
Professional Development	Calculation based on 8 indicators (range 0-100)	At least **75**% of schools implemented component with fidelity (score ≥ 78)	42%	**No**	28%	**No**	34%	**No**
On-site consultancies	Calculation based on 1 indicator (range 0-100)	At least **75**% of schools implemented component with fidelity (score ≥ 80)	91%	**Yes**	89%	**Yes**	96%	**Yes**
Classroom Planning and Instruction	Calculation based on 9 indicators (range 0-160)	At least **75**% of schools implemented component with fidelity (score ≥ 138)	47%	**No**	57%	**No**	48%	**No**
School Planning and Arts	Calculation based on 4 indicators	At least **75**% of schools implemented component with	49%	**No**	23%	**No**	18%	**No**

Programming	(range 0-40)	fidelity (score ≥ 31)						

Table 18. Overall Student Arts Achievement Impact Analysis Results (HLM)

Outcomes	Treatment Group N of Clusters	Treatment Group N of Students	Control Group N of Clusters	Control Group N of Students	Control Group Mean	Impact Estimate	Standardized Effect Size (Hedge's g)	Impact Standard Error	p-value
Art achievement Year 1	43	1,619	34	1,487	61.66	4.59	0.28	1.17	0.000
Art achievement Year 2	47	1,760	36	1,699	66.21	3.01	0.20	1.30	0.023
Art achievement Year 3	44	1,621	34	1,656	67.30	1.42	0.09	1.34	0.295

New Table 19.

Student Arts Achievement Impact Analysis Results by Arts Discipline (HLM)

Outcomes	Treatment Group N of Clusters	Treatment Group N of Students	Control Group N of Clusters	Control Group N of Students	Control Group Mean	Impact Estimate	Standardized Effect Size (Hedge's g)	Impact Standard Error	p-value
Dance achievement Year 1	11	374	9	379	64.73	2.96	0.16	2.86	0.322
Music achievement Year 1	11	430	9	492	65.35	4.24	0.28	2.01	0.048
Theater achievement Year 1	10	364	8	256	54.30	7.43	0.43	3.09	0.026
Visual Arts achievement Year 1	11	451	8	360	58.60	1.55	0.12	1.56	0.335
Dance achievement Year 2	12	437	10	518	71.36	1.53	0.10	1.47	0.307
Music achievement Year 2	12	435	9	488	64.21	7.75	0.47	2.87	0.013
Theater achievement Year 2	11	388	9	276	63.54	2.29	0.15	3.23	0.486
Visual Arts achievement Year 2	12	500	8	417	63.95	-1.24	-0.10	1.38	0.380
Dance achievement Year 3	11	398	10	446	73.23	-2.34	-0.14	2.58	0.377
Music	12	496	9	472	62.30	4.66	0.29	2.80	0.111

Outcomes	Treatment Group N of Clusters	Treatment Group N of Students	Control Group N of Clusters	Control Group N of Students	Control Group Mean	Impact Estimate	Standardized Effect Size (Hedge's g)	Impact Standard Error	p-value
achievement Year 3									
Theater achievement Year 3	9	264	8	309	69.35	-0.28	-0.02	2.31	0.907
Visual Arts achievement Year 3	12	463	7	429	65.17	-0.81	-0.07	0.91	0.398

Table 20. Overall Teacher Instructional Practice Impact Analysis Results (Linear Regression)

Outcomes	Treatment Group N of Clusters	Treatment Group N of Teachers	Control Group N of Clusters	Control Group N of Teachers	Control Group Mean	Impact Estimate	Standardized Effect Size (Hedge's g)	Impact Standard Error	p-value
Instructional practice Year 1	34	34	15	15	3.72	0.35	0.53	0.14	0.018
Instructional practice Year 2	21	21	12	12	4.06	0.38	0.59	0.18	0.044
Instructional practice Year 3	26	26	9	9	4.18	0.21	0.27	0.21	0.325

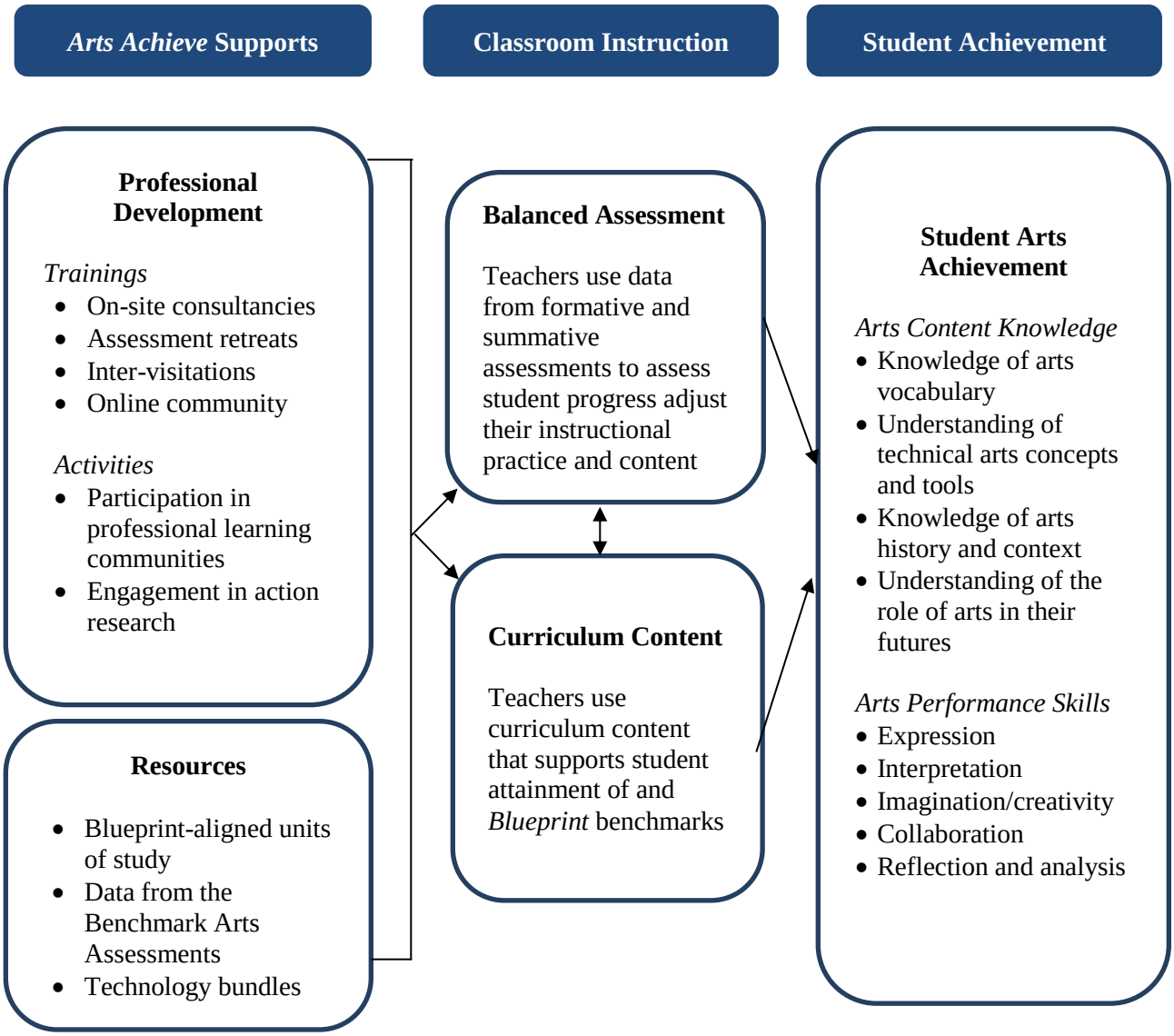

Figure 1. Arts Achieve Logic Model.

<table>
<tr><td colspan="2">

The New York City Department of Education (NYCDOE) consists of over 1,700 schools that serve 1.1 million students each year. The NYCDOE Office of Arts and Special Projects (OASP) supports universal access to arts education and increased quality in the arts through supports for school leaders, the development of curriculum and professional development for teachers of dance, music, theater, visual art, and the moving image.

</td></tr>
<tr><td>

Founded in 1979, ArtsConnection provides in-school and after school programs taught by professional teaching artists in dance, music, theater, and visual arts in NYC schools.

</td><td>

For twenty years, the 92Y Dance Education Laboratory has provided a professional development program to dance educators nationwide interested in developing their teaching practice.

</td></tr>
<tr><td>

Lead Partner

For 36 years our community of professional artists has been teaching visual arts to hundreds of thousands of New York City children in under-served public schools, daycare centers, and community-based organizations. We offer quality art workshops where children explore their creativity and learn the joys and benefits of making art. Studio also collaborates with and develops the ability of those who provide or support arts programming and creative development for youth both in and outside of schools.

</td><td>

The Weill Music Institute creates visionary programs that embody Carnegie Hall's commitment to music education. These programs inspire audiences of all ages and nurture tomorrow's musical talent, reaching more than 400,000 people each year through national and international partnerships, in New York City schools and community settings, and at Carnegie Hall.

</td></tr>
<tr><td>

Smithsonian's Cooper-Hewitt, National Design Museum is the only museum in the nation devoted exclusively to historic and contemporary design. The Museum presents compelling perspectives on the impact of design on daily life through active educational and curatorial programming.

</td><td>

Metis Associates is a national research and consulting firm headquartered in New York City. Metis has over 35 years of experience in education-based evaluation, working with a wide range of organizations committed to making a meaningful difference in the lives of children, families, and communities.

</td></tr>
</table>

Figure 2. *Arts Achieve Arts Partners.*

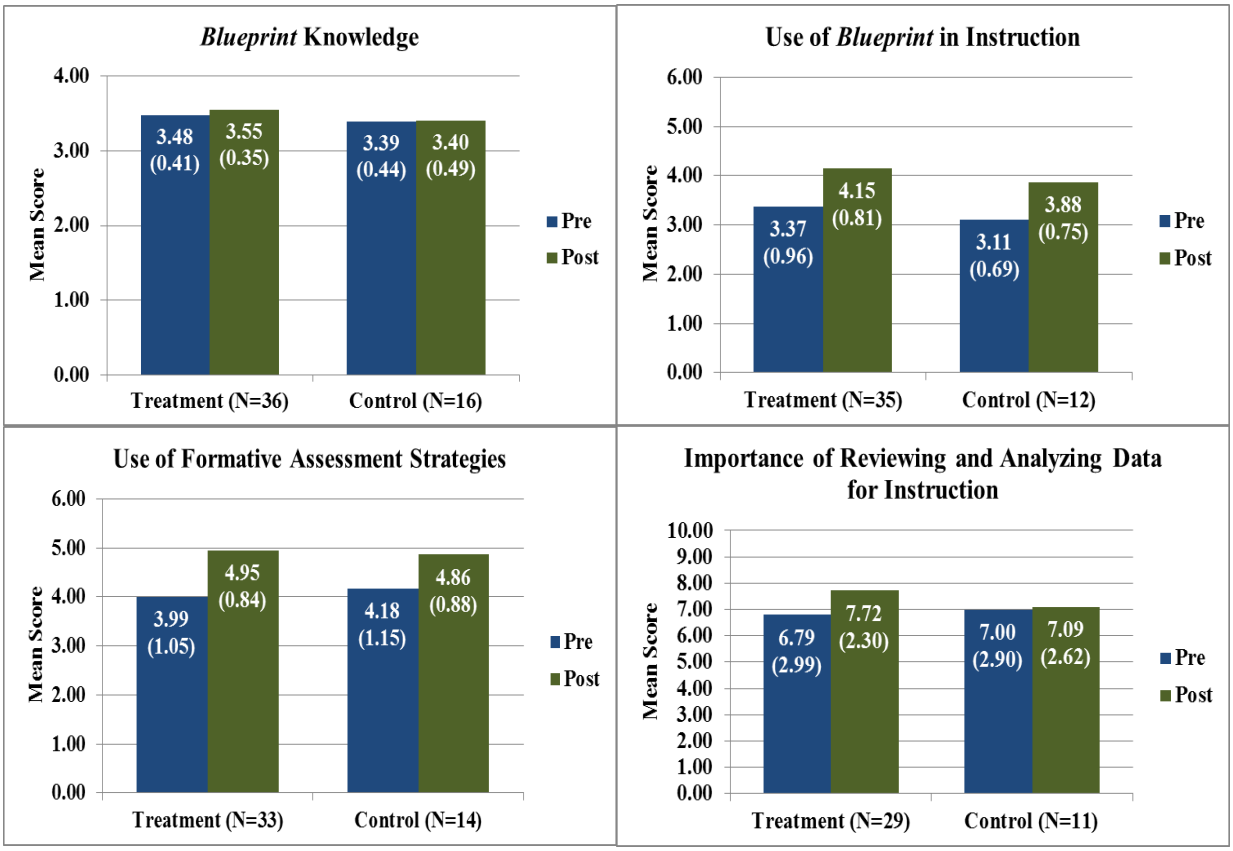

Figure 3. Arts Teachers' Perceptions of Their Content Knowledge and Instructional Practices from Pre- to Post-Survey. Note: The y-axes represent the response scale range. Standard deviations are presented in parentheses below the mean score.

Appendix

Table A1.
*Inter-rater Reliabilities of the **2011-2012** Fifth-Grade Benchmark Dance Assessments*

Task Number	Task and Description		Fall 2011		Spring 2012	
			N Cases	Kappa	N Cases	Kappa
1.1	Stylistic Hallmarks in Dance: Written interpretation of a masterwork dance performance	Content	394	0.56**	399	0.58**
1.2		Comprehension	394	0.64**	399	0.62**
1.3		Usage	393	0.62**	399	0.67**
2.1	Analysis of Dance Elements: Analyzing dance elements in a masterwork dance prompt	Body	390	0.65**	398	0.74**
2.2		Dynamics	388	0.88**	391	0.87**
2.3		Space	391	0.76**	397	0.72**
2.4		Relationship	386	0.82**	385	0.74**
3.1	Dance Styles and Genres: Comparing and contrasting dance in other styles and genres	Description of Selected Style	370	0.63**	396	0.68**
3.2		Compare and Contrast	252	0.73**	360	0.64**
4.1	Dance Performance: Performs masterwork dance prompt, choreographs solo performance, and collaborates on a duet	Movement Replication	383	0.42**	384	0.14**
4.2		Solo Performance, Creativity	379	0.40**	379	0.26**
4.3		Solo Performance, Performance Quality	379	0.40**	384	0.32**
4.4		Duet Collaboration	377	0.33**	383	0.24**
5.1	Choreographic Choices	Content	338	0.60**	340	0.75**
5.2		Comprehension	338	0.64**	341	0.71**
5.3		Usage	337	0.62**	341	0.72**

*$p < .05$, **$p < .001$.

Table A2.
*Inter-rater Reliabilities of the **2012-2013** Fifth-Grade Benchmark Dance Assessments*

Task Number	Task and Description		Fall 2012		Spring 2013	
			N Cases	Kappa	N Cases	Kappa
1.1	Stylistic Hallmarks in Dance: Interpret elements of style/genre in a masterwork dance performance	Content	503	0.64**	497	0.21**
1.2		Comprehension	503	0.65**	496	0.28**
1.3		Usage	503	0.63**	497	0.21**
2.1	Dance Analysis: Recognize, identify, and categorize dance elements in a masterwork dance prompt	Body	494	0.82**	493	0.54**
2.2		Dynamics	474	0.94**	483	0.80**
2.3		Space	475	0.88**	482	0.55**
2.4		Relationship	435	0.92**	467	0.49**
3.1	Dance Styles and Genres: Compare and contrast another dance style/genre to the masterwork prompt	Description of Selected Style	478	0.75**	474	0.30**
3.2		Compare and Contrast	384	0.69**	421	0.22**
4.1	Dance Performance: Learn a dance phrase related to the masterwork prompt, create a solo dance study using elements related to the prompt and collaborate on a duet that combines both	Movement Replication	472	0.39**	485	0.52**
4.2		Solo Performance, Creativity	471	0.46**	483	0.48**
4.3		Solo Performance, Performance Quality	472	0.43**	483	0.55**
4.4		Duet Collaboration	474	0.52**	487	0.50**
5.1	Choreographic Choices: Write speculatively and imaginatively about creating a piece of choreography	Content	405	0.71**	463	0.33**
5.2		Comprehension	405	0.72**	462	0.34**
5.3		Usage	405	0.65**	462	0.25**

*$p < .05$, **$p < .001$.

Table A3.

*Inter-rater Reliabilities of the **2013-2014** Fifth-Grade Benchmark Dance Assessments*

Task Number	Task and Description		Fall 2013		Spring 2014	
			N Cases	Kappa	N Cases	Kappa
1.1	Dance Elements: Recognize, identify, and categorize dance elements in a masterwork dance prompt	Body	465	0.52**	456	0.68**
1.2		Dynamics	458	0.78**	446	0.88**
1.3		Space	464	0.64**	452	0.79**
1.4		Relationship	445	0.60**	442	0.71**
2	Stylistic Hallmarks in Dance		462	0.46**	445	0.63**
3	Describe Dance Styles and Genres: Describe style/genre of dance		457	0.37**	442	0.40**
4	Compare and Contrast Dance Styles and Genres		428	0.41**	428	0.43**
5.1	Dance Performance: Learn a dance phrase related to the masterwork prompt, create a solo dance study using elements related to the prompt and collaborate on a duet that combines both	Movement Replication	449	0.42**	446	0.51**
5.2		Solo Performance, Creative Invention	460	0.49**	447	0.49**
5.3		Solo Performance, Performance Quality	460	0.47**	447	0.46**
5.4		Duet Collaboration	461	0.32**	448	0.28**
5.5		Responds to Varying Demands of Audience, Task, and Purpose	459	0.32**	446	0.47**
6.1	Choreographic Choices: Write speculatively and imaginatively about creating a piece of choreography	Central Themes, Ideas, Content	462	0.44**	451	0.52**
6.2		Development	461	0.49**	451	0.48**
6.3		Analysis	461	0.48**	451	0.47**
6.4		Writing Conventions	461	0.40**	451	0.47**

*$p < .05$, **$p < .001$.

Table A4.
*Inter-rater Reliabilities of the **2011-2012** Middle School Benchmark Dance Assessments*

Task Number	Task and Description		Fall 2011 N Cases	Fall 2011 Kappa	Spring 2012 N Cases	Spring 2012 Kappa
1.1	Dance Analysis: Recognize, identify, and categorize dance elements of a masterwork dance prompt	Choreographic Devices	339	1.0**	271	0.64**
1.2		Body	342	0.88**	293	0.87**
1.3		Dynamics	342	0.91**	292	0.94**
1.4		Space	342	0.92**	292	0.84**
1.5		Relationships	340	0.88**	290	0.87**
2.1	Movement Replication and Trio Performance: Collaborate on a trio of a masterwork dance prompt	Individual Accuracy of Movement	343	0.43**	285	0.45**
2.2		Individual Performance Quality	338	0.45**	285	0.55**
2.3		Group Collaboration	343	0.34**	285	0.65**
3.1	Solo Dance Performance: Choreograph and perform solo routine using elements from masterwork dance prompt	Creativity	341	0.46**	285	0.43**
3.2		Performance Quality	342	0.40**	285	0.50**
4	Choreographic Choices: Description of choreographic choices for solo		333	0.76**	284	0.82**
5	Dance Styles and Genres: Compare and contrast masterwork dance prompt to dancing in another style/genre		280	0.83**	261	0.74**
6.1	The Expression of Dance: Written response to the masterwork dance prompt and solo performance, addressing elements of dance and the relationship between movement and expression	Content	243	0.78**	234	0.81**
6.2		Comprehension	243	0.79**	234	0.76**
6.3		Usage	243	0.76**	233	0.74**

*$p < .05$, **$p < .001$.

Table A5.

*Inter-rater Reliabilities of the **2012-2013** Middle School Benchmark Dance Assessments*

Task Number	Task and Description		Fall 2012		Spring 2013	
			N Cases	Kappa	N Cases	Kappa
1.1	Dance Analysis: Recognize, identify, and categorize dance elements of a masterwork dance prompt	Choreographic Devices	315	0.84**	353	0.74**
1.2		Body	378	0.86**	388	0.59**
1.3		Dynamics	375	0.88**	386	0.77**
1.4		Space	376	0.80**	386	0.64**
1.5		Relationships	375	0.76**	384	0.53**
2.1	Movement Replication and Trio Performance: Collaborate on a trio staging of a dance phrase based on the masterwork dance prompt	Individual Accuracy of Movement	354	0.49**	384	0.52**
2.2		Individual Performance Quality	353	0.42**	385	0.51**
2.3		Group Collaboration	354	0.67**	379	0.40**
3.1	Solo Dance Performance: Choreograph and perform solo dance study inspired by the masterwork dance prompt	Creativity	341	0.51**	380	0.46**
3.2		Performance Quality	342	0.41**	379	0.51**
4	Choreographic Choices: Describe their choreographic choices for the solo dance study		361	0.77**	381	0.41**
5	Dance Styles and Genres: Compare and contrast masterwork dance prompt to dancing in another style/genre		323	0.81**	336	0.34**
6.1	The Expression of Dance: Write about the masterwork dance prompt and/or the solo dance study, addressing the relationship between movement and expression	Content	304	0.83**	329	0.34**
6.2		Comprehension	304	0.84**	329	0.36**
6.3		Usage	304	0.79**	329	0.30**

*$p < .05$, **$p < .001$.

Table A6.

*Inter-rater Reliabilities of the **2013-2014** Middle School Benchmark Dance Assessments*

Task Number	Task and Description		Fall 2013		Spring 2014	
			N Cases	Kappa	N Cases	Kappa
1.1	Dance Elements: Recognize, identify, and categorize dance elements of a masterwork dance prompt	Choreographic Devices	327	0.62**	329	0.72**
1.2		Body	376	0.43**	372	0.50**
1.3		Dynamics	368	0.59**	367	0.72**
1.4		Space	375	0.64**	368	0.61**
1.5		Relationships	373	0.51**	362	0.53*
2.1	Movement Replication and Trio Performance: Collaborate on a trio staging of a dance phrase based on the masterwork dance prompt	Individual Accuracy of Movement	277	0.63**	364	0.57**
2.2		Individual Performance Quality	277	0.63**	364	0.51**
2.3		Group Collaboration	276	0.70**	364	0.56**
3.1	Solo Dance Performance: Choreograph and perform an original solo dance study inspired by the masterwork dance prompt	Creative Invention	275	0.71**	364	0.61**
3.2		Performance Quality	276	0.66**	364	0.65**
3.3		Responds to Varying Demands of Audience, Task, and Purpose	276	0.74**	361	0.56**
4	Choreographic Choices: Describe their choreographic choices for the solo dance study		378	0.49**	355	0.50**
5	Dance Styles and Genres: Compare and contrast masterwork dance prompt to dancing in another style/genre		355	0.40**	354	0.32**
6.1	Expression of Dance: Write about the masterwork dance prompt and/or the solo dance study, addressing the relationship between movement and expression	Central Themes, Ideas, Content	314	0.39**	333	0.43**
6.2		Development	314	0.37**	332	0.40**
6.3		Analysis	314	0.35**	332	0.38**
6.4		Writing Conventions	314	0.34**	332	0.27**

*$p < .05$, **$p < .001$.

Table A7.
*Inter-rater Reliabilities of the **2011-2012** High School Benchmark Dance Assessments*

Task Number	Task and Description		Fall 2011		Spring 2012	
			N Cases	Kappa	N Cases	Kappa
1	Stage Directions: Recognize and identifying stage directions of a masterwork dance prompt		224	0.98**	230	0.98**
2.1	Dance Analysis: Recognize, identify, and analyze movements of a masterwork dance prompt	Stage Placement	229	0.85**	220	0.89**
2.2		Movement, Body, Shape, and Formation	254	0.64**	246	0.63**
2.3		Contrasting Dance Elements	253	0.65**	228	0.67**
2.4		Movement Qualities	256	0.95**	247	0.97**
3	The Body and Movement: Demonstrate basic knowledge of anatomy and kinesiology related to dance		239	0.74**	240	0.82**
4	Dance as a Profession: Demonstrate knowledge of roles in the dance profession		219	0.71**	225	0.77**
5	Dance Styles and Genre: Compare and contrast different dance styles/genres		233	0.65**	223	0.57**
6.1	Dance as Expression: Written response of masterwork dance prompt addressing the relationship between movement and expression of themes	Content	213	0.72**	220	0.67**
6.2		Comprehension	212	0.54**	220	0.68**
6.3		Usage	211	0.68**	220	0.52**
7.1	Solo Dance Performance: Participate in warm─up and create a solo routine based on a masterwork dance prompt	Accuracy of Movement	223	0.35**	221	0.34**
7.2		Creativity	223	0.34**	220	0.34**

Task Number	Task and Description		Fall 2011		Spring 2012	
			N Cases	Kappa	N Cases	Kappa
7.3		Performance Quality	224	0.34**	221	0.38**
8.1	Group Dance Performance: Collaborate to create a group performance combining solos	Accuracy of Movement	225	0.31**	194	0.47**
8.2		Choreographic Structure	225	0.05	194	0.24**
8.3		Group Collaboration	227	0.37**	195	0.13*
9.1	Reflecting on Dance Performance: Written reflection on dance performance	Relevance and Insight	206	0.63**	215	0.58**
9.2		Usage	207	0.60**	213	0.60**

*$p < .05$, **$p < .001$.

Table A8.
*Inter-rater Reliabilities of the **2012-2013** High School Benchmark Dance Assessments*

Task Number	Task and Description		Fall 2012		Spring 2013	
			N Cases	Kappa	N Cases	Kappa
1	Stage Directions: Recognize and identify stage directions of a masterwork dance prompt		332	0.92**	270	0.81**
2.1	Dance Analysis: Recall, describe, and analyze movements of a masterwork dance prompt	Stage Placement and Direction of Motion	292	0.89**	271	0.29**
2.2		Describing Dance Movements	332	0.81**	278	0.32**
2.3		Counterpoint in Choreography	333	0.79**	273	0.34**
2.4		Movement Qualities	338	0.99**	280	0.78**
3	The Body and Movement: Demonstrate basic knowledge of anatomy and kinesiology related to dance		319	0.89**	279	0.63**
4	Dance as a Profession: Demonstrate knowledge of roles in the dance profession		324	0.77**	272	0.27**
5	Dance Styles and Genre: Compare and contrast another dance style/genre to the masterwork dance prompt		315	0.72**	273	0.33**
6.1	Expression and Meaning in Dance: Write a critical response to the masterwork dance prompt, addressing how movement, choreographic, and production choices convey the work's themes	Content	314	0.75**	252	0.28**
6.2		Comprehension	315	0.72**	252	0.33**
6.3		Usage	315	0.68**	252	0.14**
7.1	Solo Dance Performance: Participate in	Accuracy of	310	0.39**	267	0.45**

Task Number	Task and Description		Fall 2012		Spring 2013	
			N Cases	Kappa	N Cases	Kappa
	warm─up and create a solo dance study inspired by a masterwork dance prompt	Movement				
7.2		Creativity	310	0.35**	267	0.44**
7.3		Performance Quality	310	0.44**	267	0.28**
8.1	Group Dance Performance: Collaborate to create a group study combining elements of the solos and themes, using choreographic devices	Accuracy of Movement	318	0.50**	271	0.45**
8.2		Choreographic Structure	317	0.34**	272	0.36**
8.3		Group Collaboration	321	0.41**	272	0.33**
9.1	Reflecting on Dance Performance: Written reflection on dance performance	Relevance and Insight	314	0.73**	267	0.20**
9.2		Usage	313	0.71**	268	0.13**

*$p < .05$, **$p < .001$.

Table A9.
*Inter-rater Reliabilities of the **2013-2014** High School Benchmark Dance Assessments*

Task Number	Task and Description		Fall 2013		Spring 2014	
			N Cases	Kappa	N Cases	Kappa
1	Stage Directions: Recognize and identify stage directions		379	0.90**	287	0.64**
2.1	Dance Analysis: Recall, describe, and analyze movements of a masterwork dance prompt	Stage Placement and Direction of Motion	358	0.69**	236	0.57**
2.2		Describing Dance Movements	372	0.32**	286	0.58**
2.3		Movement Memory	386	0.91**	254	0.87**
2.4		Movement Qualities	391	0.78**	292	0.56**
3	Anatomy and Dance: Demonstrate basic knowledge of anatomy and kinesiology related to dance		389	0.74**	292	0.79**
4	Dance Related Professions: Demonstrate knowledge of roles in the dance profession		371	0.35**	286	0.40**
5	Dance Styles and Genre: Compare and contrast another dance style/genre to the masterwork dance prompt		365	0.42**	275	0.38**
6.1	Expression and Meaning in Dance: Write a critical response to the masterwork dance prompt, addressing how movement, choreographic, and production choices convey the work's themes	Central Themes, Ideas, Content	364	0.32**	270	0.25**
6.2		Development	363	0.26**	270	0.27**
6.3		Analysis	364	0.26**	270	0.25*
6.4		Writing Conventions	364	0.39**	270	0.12*
7.1	Solo Dance Performance: Participate in warm─up and create a solo dance study	Accuracy to Task	359	0.63**	282	0.50**
7.2		Creative Invention	359	0.56**	286	0.57**

Task Number	Task and Description		Fall 2013		Spring 2014	
			N Cases	Kappa	N Cases	Kappa
7.3	inspired by a masterwork dance prompt	Performance Quality	359	0.58**	285	0.56**
8.1	Group Dance Performance: Collaborate to create a group study combining elements of the solos and themes, using choreographic devices	Accuracy to Task	353	0.54**	290	0.54**
8.2		Structuring Choreography for Expression	347	053**	290	0.58**
8.3		Group Collaboration	352	0.51**	290	0.28**
8.4		Responds to Varying Demands of Audience, Task, and Purpose	338	0.57**	285	0.48**
9.1	Dance Connections: Write a reflection on the study of dance and its connections	Relevance and Insight	356	0.33**	286	0.32**
9.2		Writing Conventions	353	0.23**	286	0.22**

*$p < .05$, **$p < .001$.

Table A10.

*Inter-rater Reliabilities of the **2011-2012** Fifth-Grade Benchmark Music Assessments*

Task Number	Task and Description		Fall 2011		Spring 2012	
			N Cases	Kappa	N Cases	Kappa
1.1	Completing a Composition: Demonstrating knowledge of music elements	Music Vocabulary	432	0.90**	402	0.93**
1.2		Reading Music I	–	–	–	–
1.3		Reading Music II	–	–	–	–
1.4		Reading Music III	–	–	–	–
1.5		Reading Music IV	432	0.93**	425	0.97**
2	Musical Form: Identifying Musical Form		431	0.89**	390	0.88**
3	Rhythmic Composition: Composing a short rhythm		53	0.69**	411	0.91**
4.1	Planning a Composition: Selecting appropriate elements for a musical composition based on a short film	Instruments	432	0.81**	423	0.83**
4.2		Dynamics	431	0.87**	356	0.91**
4.3		Tempo	432	0.85**	390	0.85**
5.1	Identifying Elements of Music: Listening to a music compositions and identifying the elements of music	Musical Elements	432	0.77**	419	0.75**
5.2		Writing Skills	432	0.75**	419	0.73**
6.1	Listening to and Comparing Musical Compositions: Listening to two music compositions and identifying elements of music	Musical Elements	431	0.83**	411	0.86**
6.2		Writing Skills	432	0.70**	399	0.78**
7.1	Music Performance: Performing learned instrumental or vocal piece (Note: Students had a choice between instrumental and vocal tasks)	Instrumental	50	1.00**	100	1.00**
7.2		Vocal	246	0.91**	280	1.00**

*$p < .05$, **$p < .001$.

Table A11.
*Inter-rater Reliabilities of the **2012-2013** Fifth-Grade Benchmark Music Assessments*

Task Number	Task and Description		Fall 2012 N Cases	Fall 2012 Kappa	Spring 2013 N Cases	Spring 2013 Kappa
1.1	Completing a Composition: Demonstrating knowledge of music elements, notation, and vocabulary	Music Vocabulary	291	0.91**	352	0.72**
1.2		Reading Music I	—	—	—	—
1.3		Reading Music II	—	—	—	—
1.4		Reading Music III	—	—	—	—
1.5		Reading Music IV	416	0.96**	412	0.92**
2	Musical Form: Identifying Musical Form		378	0.86**	379	0.87**
3	Rhythmic Composition: Composing a short rhythm		369	0.92**	383	0.88**
4.1	Planning a Composition: Selecting appropriate elements for a musical composition based on a short film	Instruments	421	0.73**	414	0.48**
4.2		Dynamics	371	0.86**	381	0.74**
4.3		Tempo	317	0.88**	344	0.67**
5.1	Listening to and Comparing Musical Compositions	Instruments	424	0.83**	418	0.73**
5.2		Dynamics	351	0.79**	374	0.69**
5.3		Tempo	368	0.86**	385	0.66**
5.4		Tone	387	0.90**	391	0.88**
6.1	Analysis of Music: Written reflection on one of the musical composition prompts	Elements of Music	414	0.62**	394	0.21**
6.2		Content	409	0.60**	394	0.26**
6.3		Comprehension	379	0.61**	394	0.26**
6.4		Usage	379	0.64**	394	0.32**
7.1	Music Performance: Performing learned instrumental or vocal piece (Note: Students had a choice between instrumental and vocal tasks)	Instrumental	92	0.99**	100	0.96**
7.2		Vocal	147	0.98**	301	1.00**

*$p < .05$, **$p < .001$.

Table A12.

*Inter-rater Reliabilities of the **2013-2014** Fifth-Grade Benchmark Music Assessments*

Task Number	Task and Description		Fall 2013		Spring 2014	
			N Cases	Kappa	N Cases	Kappa
1.1	Identifying Elements of a Music Score: Demonstrating knowledge of music elements, notation, and vocabulary	Tempo	—	—	—	—
1.2		Time Signature	—	—	—	—
1.3		Dynamics	—	—	—	—
1.4		Music Notation	—	—	—	—
1.5		Steps and Leaps	—	—	395	0.79**
2	Rhythmic Composition: Composing a short rhythmic composition		332	0.86**	375	0.81**
3.1	Identifying Elements of Music: Listening to a music composition and identifying the elements of music	Instruments	388	0.79**	399	0.76**
3.2		Dynamics	—	—	—	—
3.3		Tempo	—	—	—	—
4.1	Written Reflection of a Music Composition: Written reflection on the relationship between mood and the musical components in a composition	Mood	386	0.26**	398	0.17**
4.2		Central Themes, Ideas, Content	388	0.09*	398	0.11*
4.3		Development	388	0.06	398	0.17**
4.4		Analysis	388	0.11*	398	0.11*
4.5		Writing Conventions	386	0.23**	398	0.32**
5.1	Music Performance: Performing learned vocal piece in groups	Individual Accuracy	389	0.42**	395	0.54**
5.2		Ensemble Participation	384	0.33**	394	0.33**
5.3		Responds to Varying Demands of Audience, Task, and Purpose	380	0.43**	392	0.38**

*$p < .05$, **$p < .001$.

Table A13.
*Inter-rater Reliabilities of the **2011-2012** Middle School Benchmark Music Assessments*

Task Number	Task and Description		Fall 2011 N Cases	Fall 2011 Kappa	Spring 2012 N Cases	Spring 2012 Kappa
1.1	Elements of Music: Listening to a music composition and identifying the elements of music	Tempo	264	0.76**	416	0.90**
1.2		Dynamics	264	0.73**	394	0.85**
1.3		Blending of Voices	264	0.54**	424	0.78**
1.4		Instruments/Voices	264	0.47**	416	0.79**
1.5		Artistic Choices	264	0.57**	408	0.75**
2	Composition: Composing a short rhythm		443	0.63**	422	0.94**
3	Note Identification: Identifying notes of a composition		264	0.92**	415	0.98**
4A.1	Identifying the Elements of Music ─ Recording A: Listening to two different music compositions and identifying the elements of music	Instruments/Voices	443	0.64**	423	0.83**
4A.2		Melody	442	0.65**	414	0.86**
4A.3		Harmony	443	0.67**	401	0.87**
4A.4		Mood	443	0.55**	418	0.76**
4A.5		Tempo	443	0.79**	402	0.87**
4B.1	Identifying the Elements of Music ─ Recording B: Listening to two different music compositions and identifying the elements of music	Instruments/Voices	443	0.60**	422	0.81**
4B.2		Melody	443	0.65**	413	0.86**
4B.3		Harmony	443	0.69**	398	0.86**
4B.4		Mood	443	0.53**	418	0.77**
4B.5		Tempo	443	0.77**	403	0.90**
5.1	Genres and Styles: Comparing and contrasting two different music compositions	Similarities	442	0.55**	422	0.82**
5.2		Differences	442	0.60**	423	0.83**
5.3		Writing Skills	442	0.56**	422	0.76**
6	Reading Music: Listening to a music composition and identifying the written score		383	0.91**	421	0.98**
7	Musical Connections: Identifying and describing roles in music		443	0.79**	376	0.99**
8	Music Performance: Performing composed rhythm		─	─	414	0.77**

*$p < .05$, **$p < .001$.

Table A14.
*Inter-rater Reliabilities of the **2012-2013** Middle School Benchmark Music Assessments*

Task Number	Task and Description		Fall 2012		Spring 2013	
			N Cases	Kappa	N Cases	Kappa
1.1	Elements of Music: Listening to a music composition and identifying the elements of music	Tempo	426	0.87**	393	0.83**
1.2		Dynamics	421	0.88**	387	0.81**
1.3		Composer's Choices	421	0.68**	393	0.62**
2	Composition: Composing a short rhythm		427	0.95**	402	0.78**
3	Note Identification: Identifying notes of a composition		424	0.97**	400	0.89**
4A.1	Genres and Styles ─ Elements of Music: Listening to two different music compositions and identifying the elements of music	Instruments/Voices	428	0.72**	404	0.66**
4A.2		Tempo	424	0.87**	400	0.68**
4A.3		Mood	429	0.72**	402	0.56**
4B	Genres and Styles ─ Similarities and Differences: Comparing the elements of music of the musical composition prompts		419	0.73**	403	0.52**
4C.1	Genres and Styles ─ Written Reflection: Written reflection on the musical composition prompts in relation to a scene	Writing Skills	428	0.61*	─	─
4C.2		Content	─	─	404	0.53**
4C.3		Comprehension	─	─	404	0.46**
4C.4		Usage	─	─	404	0.42**
5	Reading Music: Listening to a music composition and identifying the written score		428	0.97**	392	0.91**
6	Music Performance: Performing composed rhythm		389	0.94**	─	─
7	Musical Connections: Identifying and describing roles in music		425	0.80**	394	0.67**

*$p < .05$, **$p < .001$.

Table A15.
*Inter-rater Reliabilities of the **2013-2014** Middle School Benchmark Music Assessments*

Task Number	Task and Description		Fall 2013		Spring 2014	
			N Cases	Kappa	N Cases	Kappa
1.1	Identifying Elements of Music Score: Demonstrating knowledge of music elements, notation, and vocabulary	Tempo	—	—	—	—
1.2		Time Signature	—	—	—	—
1.3		Dynamics	—	—	—	—
1.4		Music Notation	—	—	—	—
1.5		Note Identification	—	—	362	0.92**
2	Composition: Composing a short rhythm		412	0.76**	390	0.85**
3A.1	Identifying Elements of Music: Listening to two different musical compositions and identifying mood and the elements of music	Composition #1:Voices/Instruments	412	0.71**	388	0.74**
3A.2		Composition #1:Tempo	—	—		
3A.3		Composition #1:Dynamics	—	—		
3A.4		Composition #1:Mood	390	0.36**	379	0.24**
3B.1		Composition #2:Voices/Instruments	412	0.56**	387	0.50**
3B.2		Composition #2:Tempo	—	—	—	—
3B.3		Composition #2:Dynamics	—	—	—	—
3B.4		Composition #2:Mood	386	0.23**	372	0.29**
4.1	Compare and Contrast: Identifying the similarities and differences in two musical compositions	Similarities	—	—	—	—
4.2		Differences	—	—	—	—
5.1	Written Reflection of a Music Composition: Written reflection on the student's preferred music composition using the elements of music	Central Themes, Ideas, Content	406	0.10*	383	0.16**
5.2		Development	406	0.00	383	0.30**
5.3		Analysis	404	0.04	384	0.29**
5.4		Writing Conventions	405	0.22**	384	0.21**
6	Reading Music: Listening to a musical composition while identifying the written score		398	0.84**	382	0.92**
7.1	Music Performance: Performing composed rhythm	Individual Accuracy	383	0.44**	359	0.49**
7.2		Ensemble Participation	383	0.33**	357	0.42**
7.3		Responds to Varying Demands of Audience, Task, and Purpose	383	0.15**	330	0.40**
8	Musical Connections: Identifying and describing roles in music		391	0.31**	374	0.12**

*$p < .05$, **$p < .001$.

Table A16.
*Inter-rater Reliabilities of the **2011-2012** High School Benchmark Music Assessments*

Task Number	Task and Description		Fall 2011 N Cases	Fall 2011 Kappa	Spring 2012 N Cases	Spring 2012 Kappa
1	Notation, Rhythm, and Meter: Demonstrating knowledge of the basic elements of music composition		331	0.38**	356	0.91**
2	Pitch: Identifying the pitch of music notation		342	0.70**	360	0.95**
3.1	Dynamics and Tempo: Demonstrating knowledge of dynamics and tempo	Dynamics	341	0.60**	365	0.96**
3.2		Tempo	339	0.70**	365	0.98**
4	Key: Identifying the key of a musical composition		339	0.77**	367	0.99**
5A.1	Identifying the Elements of Music – Composition 1: Listening to and identifying elements of musical composition	Instruments/Voices	339	0.61**	367	0.79**
5A.2		Mood	339	0.50**	356	0.80**
5A.3		Meter	304	0.76**	311	0.96**
5A.4		Tempo	339	0.71**	330	0.96**
5A.5		Melody	339	0.70**	343	0.88**
5A.6		Dynamics	339	0.73**	293	0.98**
5B.1	Identifying the Elements of Music – Composition 2: Listening to and identifying elements of musical composition	Instruments/Voices	339	0.69**	365	0.73**
5B.2		Mood	298	0.65**	336	0.79**
5B.3		Meter	337	0.76**	297	0.94**
5B.4		Tempo	339	0.79**	325	0.94**
5B.5		Melody	339	0.75**	336	0.81**
5B.6		Dynamics	339	0.71**	292	0.89**
5C.1	Identifying the Elements of Music – Composition 3: Listening to and identifying elements of musical composition	Instruments/Voices	338	0.66**	359	0.64**
5C.2		Mood	304	0.64**	344	0.82**
5C.3		Meter	338	0.81**	295	0.96**
5C.4		Tempo	339	0.78**	325	0.92**
5C.5		Melody	338	0.73**	326	0.70**
5C.6		Dynamics	339	0.79**	297	0.91**
6.1	Genres and Styles of Music: Comparing the elements of music of the three musical composition prompts	Similarities	244	0.82**	360	0.83**
6.2		Differences	244	0.78**	356	0.85**
7	Writing a Review: Written reflection on one of the three musical compositions		244	0.87**	334	0.66**

*$p < .05$, **$p < .001$.

Table A17.
*Inter-rater Reliabilities of the **2012-2013** High School Benchmark Music Assessments*

Task Number	Task and Description		Fall 2012 N Cases	Fall 2012 Kappa	Spring 2013 N Cases	Spring 2013 Kappa
1	Notation, Rhythm, and Meter: Demonstrating knowledge of the basic elements of music composition		429	0.89**	395	0.61**
2	Pitch: Identifying the pitch of music notation		434	0.95**	395	0.92**
3.1	Dynamics and Tempo: Demonstrating	Dynamics	507	0.90**	417	0.85**
3.2	knowledge of dynamics and tempo	Tempo	506	0.93**	419	0.80**
4	Key: Identifying the key of a musical composition		516	0.98**	416	0.94**
5A.1	Identifying the Elements of Music ─	Instruments/Voices	492	0.82**	393	0.35**
5A.2	Composition 1: Listening to and	Tempo	424	0.88**	369	0.64**
5A.3	identifying elements of musical	Dynamics	345	0.91**	339	0.65**
5A.4	composition	Mood	451	0.77**	375	0.29**
5B.1	Identifying the Elements of Music ─	Instruments/Voices	458	0.80**	383	0.48**
5B.2	Composition 2: Listening to and	Tempo	399	0.90**	355	0.67**
5B.3	identifying elements of musical	Dynamics	331	0.86**	328	0.54**
5B.4	composition	Mood	417	0.78**	358	0.28**
5C.1	Identifying the Elements of Music ─	Instruments/Voices	433	0.82**	358	0.28**
5C.2	Composition 3: Listening to and	Tempo	380	0.88**	339	0.76**
5C.3	identifying elements of musical	Dynamics	311	0.92**	316	0.72**
5C.4	composition	Mood	384	0.74**	342	0.29**
6.1	Genres and Styles of Music: Comparing the elements of music of the musical composition prompts	Similarities	420	0.72**	355	0.25**
6.2		Differences	414	0.66**	351	0.24**
7.1	Writing a Review: Written reflection on one of the three musical compositions	Content	377	0.63**	341	0.21**
7.2		Comprehension	334	0.59**	341	0.20**
7.3		Usage	335	0.67**	341	0.19**

*$p < .05$, **$p < .001$.

Table A18.
*Inter-rater Reliabilities of the **2013-2014** High School Benchmark Music Assessments*

Task Number	Task and Description		Fall 2013		Spring 2014	
			N Cases	Kappa	N Cases	Kappa
1.1	Identifying Elements of Music Score: Demonstrating knowledge of music elements, notation, and vocabulary	Time Signature	—	—	—	—
1.2		Tempo	—	—	—	—
1.3		Key Signature	—	—	—	—
1.4		Crescendo	—	—	—	—
1.5		Note Identification	—	—	342	0.86**
2	Composition: Composing a short rhythmic composition, using accurate music notation		420	0.59**	370	0.82**
3	Tonality: Identifying the tonality of a musical composition		470	0.86**	375	0.94**
4A.1	Identifying the Elements of Music — Composition 1: Listening to a music composition and identifying the elements of music	Instruments/Voices	447	0.67**	371	0.54**
4A.2		Tempo	—	—	—	—
4A.3		Dynamics	—	—	—	—
4A.4		Mood	385	0.38**	347	0.36**
4B.1	Identifying the Elements of Music — Composition 2: Listening to a music composition and identifying the elements of music	Instruments/Voices	418	0.62**	367	0.66**
4B.2		Tempo	—	—	—	—
4B.3		Dynamics	—	—	—	—
4B.4		Mood	350	0.27**	345	0.34**
5.1	Compare and Contrast: Identifying the similarities and differences in two music compositions	Similarities	—	—	—	—
5.2		Differences	—	—	—	—
6.1	Music Performance: Reading and performing a composed rhythm in groups	Individual Accuracy	377	0.40**	346	0.32**
6.2		Ensemble Participation	371	0.37**	343	0.29**
6.3		Responds to Varying Demands of Audience, Task, and Purpose	370	0.42**	342	0.29**
7.1	Writing a Review: Writing a review of a musical composition and describing the relationship between the composition's mood and the elements of music	Central Themes, Ideas, Content	295	0.20**	331	0.29**
7.2		Development	295	0.29**	331	0.34**
7.3		Analysis	295	0.22**	331	0.28**
7.4		Writing Conventions	295	0.26**	331	0.28**

*$p < .05$, **$p < .001$.

Table A19.
*Inter-rater Reliabilities of the **2011-2012** Fifth-Grade School Benchmark Theater Assessments*

Task Number	Task and Description		Fall 2011		Spring 2012	
			N Cases	Kappa	N Cases	Kappa
1.1	Theater Content and Vocabulary: Demonstrating knowledge of theater content and vocabulary	Theater Vocabulary I	—	—	—	—
1.2		Theater Vocabulary II	—	—	—	—
1.3		Theater Vocabulary III	—	—	—	—
1.4		Theater Vocabulary IV	—	—	—	—
2A.1	Theater Analysis ─ Playwriting ─ Writing a short dialogue using a photograph of a play scene (Note: Students had a choice between playwriting and costume design tasks)	Development of Character	150	0.65**	207	0.56**
2A.2		Engaging Dialogue	150	0.63**	207	0.60**
2A.3		Clear Conflict	150	0.64**	207	0.55**
2B.1	Theater Analysis ─ Costume Design: Describing the relationship between the costumes and characters within a photograph of a play (Note: Students had a choice between playwriting and costume design tasks)	Description of Costume	180	0.53**	143	0.56**
2B.2		Analysis of Character	180	0.50**	143	0.62**
2B.3		Relationship Between Character and Costume	180	0.55**	143	0.54**
2B.4		Writing Skills	149	0.56**	143	0.59**
3.1	Theater Performance: Acting in a two─person scripted scene	Focus and Commitment	273	0.33**	334	0.35**
3.2		Vocal Skills	273	0.38**	332	0.31**
3.3		Physicality	270	0.48**	331	0.31**
3.4		Objectives	269	0.33**	330	0.42**
4.1	Group Tableau: Group tableau performance	Reacts Physically to Imaginary Circumstances	—	—	—	—
4.2		Responds Physically to Other Characters	—	—	—	—
4.3		Reveals the Character through Physical Presence	—	—	—	—
4.4		Reveals the Setting through Movement	—	—	—	—

*$p < .05$, **$p < .001$.

Table A20.
*Inter-rater Reliabilities of the **2012-2013** Fifth-Grade School Benchmark Theater Assessments*

Task Number	Task and Description		Fall 2012		Spring 2013	
			N Cases	Kappa	N Cases	Kappa
1.1	Theater Content and Vocabulary: Demonstrating knowledge of theater content and vocabulary	Theater Vocabulary I	—	—	—	—
1.2		Theater Vocabulary II	—	—	—	—
1.3		Theater Vocabulary III	—	—	—	—
1.4		Theater Vocabulary IV	—	—	—	—
2A.1	Theater Analysis — Playwriting: Writing a short dialogue using a photograph of a play scene (Note: Students had a choice between playwriting and costume design tasks)	Development of Character	239	0.60**	276	0.30**
2A.2		Engaging Dialogue	237	0.65**	276	0.30**
2A.3		Clear Conflict	235	0.53**	276	0.33**
2B.1	Theater Analysis — Costume Design: Describing the relationship between the costumes and characters within a photograph of a play (Note: Students had a choice between playwriting and costume design tasks)	Description of Costume	186	0.49**	145	0.43**
2B.2		Analysis of Costume	187	0.60**	145	0.41**
2B.3		Analysis of Costume & Scene	187	0.51**	145	0.42**
2B.4		Content	187	0.59**	145	0.45**
2B.5		Comprehension	186	0.56**	145	0.41**
2B.6		Usage	188	0.67**	145	0.46**
3.1	Theater Performance: Acting in a two—person scripted scene	Focus and Commitment	425	0.34**	415	0.30**
3.2		Vocal Skills	426	0.40**	415	0.37**
3.3		Physicality	423	0.33**	415	0.42**
3.4		Objectives	425	0.39**	414	0.38**
4.1	Group Tableau: Group tableau performance	Reacts Physically to Imaginary Circumstances	—	—	—	—
4.2		Responds Physically to Other Characters	—	—	—	—
4.3		Reveals the Character through Physical Presence	—	—	—	—
4.4		Reveals the Setting through Movement	—	—	—	—

*$p < .05$, **$p < .001$.

Table A21.

*Inter-rater Reliabilities of the **2013-2014** Fifth-Grade School Benchmark Theater Assessments*

Task Number	Task and Description		Fall 2013		Spring 2014	
			N Cases	Kappa	N Cases	Kappa
1.1	Theater Content and Vocabulary: Demonstrating knowledge of theater content and vocabulary	Stage Directions	—	—	—	—
1.2		Levels	—	—	—	—
1.3		Focus	—	—	—	—
1.4		Setting	—	—	—	—
2.1	Theater Analysis — Playwriting: Writing a short dialogue using a photograph of a play scene	Central Themes, Ideas, Content-Characters	455	0.32**	435	0.36**
2.2		Central Themes, Ideas, Content — Relationships	455	0.31**	435	0.40**
2.3		Central Themes, Ideas, Content — Conflict	455	0.36**	435	0.35**
2.4		Development	455	0.41**	435	0.36**
2.5		Analysis	453	0.40**	436	0.40**
2.6		Writing Conventions	452	0.36**	436	0.43**
3.1	Theater Analysis — Costume Design: Describing the relationship between the costumes and characters within a photograph of a play	Central Themes, Ideas, Content	426	0.34**	428	0.36**
3.2		Development	426	0.28**	430	0.30**
3.3		Analysis	426	0.33**	430	0.30**
3.4		Writing Conventions	426	0.37**	430	0.34**
4.1	Theater Performance: Acting in a two—person scripted scene	Focus and Commitment	426	0.40**	418	0.40**
4.2		Diction and Projection	426	0.38**	418	0.42**
4.3		Vocal Expression	426	0.38**	418	0.42**
4.4		Physicality	423	0.36**	418	0.34**
4.5		Objectives	422	0.37**	417	0.40**
4.6		Responds to Varying Demands of Audience, Task, and Purpose	257	0.35**	412	0.40**
5.1	Group Tableau: Group tableau performance	Reacts Physically to Imaginary Circumstances	—	—	—	—
5.2		Responds Physically to Other Characters	—	—	—	—
5.3		Reveals the Character through Physical Presence	—	—	—	—
5.4		Reveals the Setting through Movement	—	—	—	—

*$p < .05$, **$p < .001$.

Table A22.

*Inter-rater Reliabilities of the **2011-2012** Middle School Benchmark Theater Assessments*

Task Number	Task and Description		Fall 2011		Spring 2012	
			N Cases	Kappa	N Cases	Kappa
1A.1	Theater Performance ─ Acting: Acting in a two-person scene and incorporating director's adjustments (Note: Students had a choice between acting, musical theater, and playwriting tasks)	Collaboration	120	0.33**	106	0.27**
1A.2		Vocal Skills	121	0.32**	106	0.35**
1A.3		Physicality	120	0.34**	106	0.47**
1A.4		Staging	121	0.27**	106	0.22**
1A.5		Objectives	119	0.27**	105	0.37**
1A.6		Responds to Direction	69	0.39**	53	0.45**
1B.1	Theater Performance ─ Musical Theater: Choreographing, rehearsing, and performing a musical number in a group and then incorporating director's adjustments (Note: Students had a choice between acting, musical theater, and playwriting tasks)	Collaboration	44	0.57**	50	0.13
1B.2		Vocal Skills	44	0.36**	49	0.65**
1B.3		Staging	44	─0.19*	49	0.36**
1B.4		Responds to Direction	30	0.72**	32	0.68**
1C.1	Theater Performance ─ Playwriting: Collaborating with a partner on completing a scripted scene and then incorporating director's adjustments (Note: Students had a choice between acting, musical theater, and playwriting tasks)	Objectives	46	0.65**	83	0.77**
1C.2		Dramatic Structure	46	0.49**	83	0.49**
1C.3		Collaboration	46	0.55**	82	0.68**
1C.4		Responds to Direction	29	0.39**	45	0.76**
2A.1	Scene Analysis ─ Acting: Watching a filmed performance and providing an analysis of the actors in the scene (Note: Students had a choice between acting and design tasks)	Analysis of Character	173	0.48**	159	0.62**
2A.2		Analysis of Relationships	171	0.48**	158	0.70**
2A.3		Writing Skills	171	0.51**	161	0.63**
2B.1	Scene Analysis ─ Design: Watching a filmed performance and providing an analysis of the playwright's intent (Note: Students had a choice between acting and design tasks)	Analysis of Intent	37	0.65**	13	0.11
2B.2		Justification of Artistic Interpretation	36	0.44*	13	0.59*
2B.3		Writing Skills	38	0.64**	12	0.82*

*$p < .05$, **$p < .001$.

Table A23.
*Inter-rater Reliabilities of the **2012-2013** Middle School Benchmark Theater Assessments*

			Fall 2012		Spring 2013	
Task Number	Task and Description		N Cases	Kappa	N Cases	Kappa
1A.1	Group Tableau #1:	Strong Physical Choices	—	—	—	—
1A.2	Group tableau	Incorporates Levels	—	—	—	—
1A.3	performance	Collaborates to Create an Image	—	—	—	—
1A.4		Responds to Prompts	—	—	—	—
1B.1	Group Tableau #2:	Strong Physical Choices	—	—	—	—
1B.2	Group tableau	Incorporates Levels	—	—	—	—
1B.3	performance	Collaborates to Create an Image	—	—	—	—
1B.4		Responds to Prompts	—	—	—	—
1C.1	Group Tableau #3:	Strong Physical Choices	—	—	—	—
1C.2	Group tableau	Incorporates Levels	—	—	—	—
1C.3	performance	Collaborates to Create an Image	—	—	—	—
1C.4		Responds to Prompts	—	—	—	—
2A.1	Theater	Collaboration	176	0.22**	139	0.46**
2A.2	Performance —	Vocal Skills	177	0.41**	139	0.47**
2A.3	Acting: Acting in a	Physicality	176	0.35**	139	0.45**
2A.4	two-person	Staging	177	0.27**	139	0.32**
2A.5	scripted scene and incorporating director's adjustments (Note:	Objectives	177	0.28**	139	0.52**
2A.6	Students had a choice between acting and playwriting tasks)	Responds to Direction	160	0.36**	132	0.43**
2B.1	Theater	Objectives	83	0.43**	104	0.30**
2B.2	Performance —	Dramatic Structure	84	0.36**	104	0.33**
2B.3	Playwriting: Collaborating with a partner on completing a scripted scene and then incorporating	Collaboration	84	0.51**	104	0.17*
2B.4	director's adjustments (Note: Students had a choice between acting and playwriting tasks)	Responds to Direction	68	0.40**	92	0.13*
3A.1	Scene Analysis —	Directorial Point of View	185	0.50**	121	0.39**
3A.2	Staging: Watching	Relationships Between Characters	187	0.57**	121	0.47**
3A.3	a filmed	Scene Analysis	177	0.54**	121	0.45**
3A.4	performance and	Content	187	0.54**	121	0.37**
3A.5	providing a written analysis of the staging (Note:	Comprehension	186	0.61**	121	0.37**
3A.6	Students had a choice between staging and design	Usage	182	0.50**	121	0.53**

Task Number	Task and Description		Fall 2012 N Cases	Fall 2012 Kappa	Spring 2013 N Cases	Spring 2013 Kappa
	tasks)					
3B.1	Scene Analysis ─	Description of Design	85	0.51**	92	0.48**
3B.2	Design: Watching	Intent of Design	83	0.45**	92	0.40**
3B.3	a filmed	Intent of Costumes	82	0.40**	92	0.42**
3B.4	performance and	Content	82	0.56**	92	0.26**
3B.5	providing a written analysis of the design (Note:	Comprehension	82	0.54**	92	0.21*
3B.6	Students had a choice between staging and design tasks)	Usage	81	0.46**	92	0.42**

*$p < .05$, **$p < .001$.

Table A24.
*Inter-rater Reliabilities of the **2013-2014** Middle School Benchmark Theater Assessments*

Task Number	Task and Description		Fall 2013		Spring 2014	
			N Cases	Kappa	N Cases	Kappa
1A.1	Group Tableau #1: Group tableau performance	Strong Physical Choices	—	—	—	—
1A.2		Incorporates Levels	—	—	—	—
1A.3		Collaborates to Create an Image	—	—	—	—
1A.4		Responds to Prompts	—	—	—	—
1B.1	Group Tableau #2: Group tableau performance	Strong Physical Choices	—	—	—	—
1B.2		Incorporates Levels	—	—	—	—
1B.3		Collaborates to Create an Image	—	—	—	—
1B.4		Responds to Prompts	—	—	—	—
1C.1	Group Tableau #3: Group tableau performance	Strong Physical Choices	—	—	—	—
1C.2		Incorporates Levels	—	—	—	—
1C.3		Collaborates to Create an Image	—	—	—	—
1C.4		Responds to Prompts	—	—	—	—
2A.1	Theater Performance — Acting: Acting in a two-person scripted scene and incorporating director's adjustments (Note: Students had a choice between acting and playwriting tasks)	Collaboration	133	0.16**	133	0.38**
2A.2		Vocal Skills	133	0.38**	133	0.47**
2A.3		Physicality	133	0.48**	132	0.34**
2A.4		Staging	133	0.52**	133	0.36**
2A.5		Objectives	133	0.32**	133	0.38**
2A.6		Responds to Varying Demands of Audience, Task, and Purpose	91	0.42**	133	0.29**
2A.7		Responds to Direction	131	0.36**	132	0.34**
2B.1	Theater Performance — Playwriting: Collaborating with a partner on completing a scripted scene and then incorporating director's adjustments (Note: Students had a choice between acting and playwriting tasks)	Central Themes, Ideas, Content — Objectives	106	0.20*	96	0.50**
2B.2		Central Themes, Ideas, Content — Structure	107	0.27**	96	0.58**
2B.3		Central Themes, Ideas, Content — Collaboration	107	0.13*	96	0.49**
2B.4		Development	107	0.22*	96	0.47**
2B.5		Analysis	107	0.20*	96	0.41**
2B.6		Writing Conventions	107	0.20*	95	0.46**
2B.7		Responds to Directions	98	0.27**	94	0.27**
3	Theater Content and Vocabulary: Demonstrating knowledge of theater content and vocabulary		—	—	—	—
4A.1	Scene Analysis — Staging: Watching a	Central Themes,	140	0.47**	148	0.46**

Task Number	Task and Description		Fall 2013		Spring 2014	
			N Cases	Kappa	N Cases	Kappa
4A.2	filmed performance and providing a written analysis of the staging (Note: Students had a choice between staging and design tasks)	Ideas, Content Development	141	0.41**	148	0.57**
4A.3		Analysis	141	0.33**	148	0.52**
4A.4		Writing Conventions	141	0.38**	148	0.54**
4B.1	Scene Analysis ─ Design: Watching a filmed performance and providing a written analysis of the design (Note: Students had a choice between staging and design tasks)	Central Themes, Ideas, Content Development	89	0.31**	59	0.39**
4B.2			89	0.30**	59	0.39**
4B.3		Analysis	89	0.22**	59	0.40**
4B.4		Writing Conventions	89	0.24*	59	0.35**

*$p < .05$, **$p < .001$.

Table A25.
*Inter-rater Reliabilities of the **2011-2012** High School Benchmark Theater Assessments*

Task Number	Task and Description		Fall 2011		Spring 2012	
			N Cases	Kappa	N Cases	Kappa
1.1	Theater Performance ─ Acting: Collaborating with a partner to complete the dialogue of a scripted scene and then performing the scene	Collaboration	229	0.27**	178	0.59**
1.2		Vocal Skills	227	0.34**	178	0.43**
1.3		Physicality	228	0.35**	178	0.29**
1.4		Staging	228	0.38**	178	0.41**
1.5		Objectives	228	0.35**	178	0.35**
1.6		Playmaking Structure	227	0.36**	177	0.30**
2A.1	Scene Analysis ─ Directing: Written Analysis of a scene in relation to the actors (Note: Students had a choice between directing and costume design tasks)	Directorial Point of View	92	0.77**	135	0.65**
2A.2		Relationship Between Characters	92	0.81**	135	0.69**
2A.3		Scene Analysis	92	0.76**	135	0.71**
2A.4		Writing Skills	91	0.66**	135	0.68**
2B.1	Scene Analysis ─ Costume Design: Written Analysis of a scene in relation to the costume design (Note: Students had a choice between directing and costume design tasks)	Description of set and costume	118	0.72**	40	0.69**
2B.2		Relationship between character and costume	118	0.76**	40	0.79**
2B.3		Description of set and costumes for change in setting	118	0.63**	40	0.51**
2B.4		Design Analysis	119	0.74**	40	0.63**
2B.5		Writing Skills	119	0.71**	40	0.49**

*$p < .05$, **$p < .001$.

Table A26.

*Inter-rater Reliabilities of the **2012-2013** High School Benchmark Theater Assessments*

Task Number	Task and Description		Fall 2012		Spring 2013	
			N Cases	Kappa	N Cases	Kappa
1.1	Theater Performance ─ Acting: Collaborating with a partner to complete the dialogue of a scripted scene and then performing the scene	Collaboration	175	0.38**	133	0.43**
1.2		Vocal Skills	176	0.32**	135	0.35**
1.3		Physicality	176	0.38**	132	0.35**
1.4		Staging	176	0.41**	133	0.46**
1.5		Objectives	175	0.44**	132	0.38**
1.6		Playmaking Structure	176	0.47**	132	0.22**
2A.1	Scene Analysis ─ Acting: Watching a filmed performance and providing an analysis of the actors in the scene (Note: Students had a choice between the acting and design tasks)	Analysis of Character	162	0.47**	82	0.57**
2A.2		Analysis of Relationships	161	0.45**	82	0.44**
2A.3		Content	162	0.41**	81	0.38**
2A.4		Comprehension	161	0.54**	81	0.46**
2A.5		Usage	162	0.53**	81	0.39**
2B.1	Scene Analysis ─ Design: Watching a filmed performance and providing an analysis of the playwright's intent (Note: Students had a choice between the acting and design tasks)	Analysis of Design Choices	15	0.29*	50	0.67**
2B.2		Analysis of Design and Character	15	0.42*	50	0.67**
2B.3		Content	15	0.44*	50	0.64**
2B.4		Comprehension	15	0.38*	50	0.49**
2B.5		Usage	15	0.53*	50	0.37*

*$p < .05$, **$p < .001$.

Table A27.

*Inter-rater Reliabilities of the **2013-2014** High School Benchmark Theater Assessments*

Task Number	Task and Description		Fall 2013		Spring 2014	
			N Cases	Kappa	N Cases	Kappa
1.1	Theater Performance ─ Acting: Collaborating with a partner to complete the dialogue of a scripted scene and then performing the scene	Collaboration	158	0.41**	116	0.21*
1.2		Vocal Skills	157	0.39**	115	0.39**
1.3		Physicality	158	0.39**	116	0.36**
1.4		Staging	158	0.47**	116	0.40**
1.5		Objectives	158	0.51**	116	0.32**
1.6		Playmaking Structure	158	0.50**	116	0.33**
1.7		Responds to Varying Demands of Audience, Task, and Purpose	158	0.60**	116	0.38**
2	Theater Content and Vocabulary: Demonstrating knowledge of theater content and vocabulary		─	─	─	─
3A.1	Scene Analysis ─ Acting: Watching a filmed performance and providing an analysis of the actors in the scene (Note: Students had a choice between the acting and staging tasks)	Central Themes, Ideas, Content	75	0.49**	76	0.65**
3A.2		Development	75	0.49**	76	0.66**
3A.3		Analysis	75	0.46**	76	0.56**
3A.4		Writing Conventions	75	0.47**	75	0.55**
3B.1	Scene Analysis ─ Staging: Watching a filmed performance and providing an analysis staging choices (Note: Students had a choice between the acting and staging tasks)	Central Themes, Ideas, Content	57	0.43**	38	0.55**
3B.2		Development	57	0.37**	38	0.46**
3B.3		Analysis	57	0.28**	38	0.58**
3B.4		Writing Conventions	57	0.50**	38	0.24**

*$p < .05$, **$p < .001$.

Table A28.
*Inter-rater Reliabilities of the **2011-2012** Fifth-Grade Benchmark Visual Arts Assessments*

Task Number	Task and Description		Fall 2011		Spring 2012	
			N Cases	Kappa	N Cases	Kappa
1.1	Elements of Art ─ Drawing Lines: Demonstrating knowledge of a variety of lines		─	─	─	─
2.1	Elements of Art ─ Visual Textures: Demonstrating knowledge of visual textures	Identifying Visual Textures	─	─	─	─
2.2		Creating Visual Textures	─	─	─	─
3.1	Drawing: Using lines and textures to create a drawing reflecting imaginative capacities	Achieves Expressive Quality	586	0.54**	577	0.65**
3.2		Uses a Variety of Lines	587	0.41**	577	0.45**
3.3		Uses a Variety of Texture	588	0.39**	576	0.51**
3.4		Uses Space Appropriately	588	0.42**	577	0.57**
4.1	Color Theory: Demonstrating knowledge of categories of color	Primary	─	─	─	─
4.2		Secondary	─	─	─	─
5.1	Tints and Shades: Creating tints and shades	Tints	─	─	─	─
5.2		Shades	─	─	─	─
6	Thinking Like an Artist ─ Elements of Art: Written response, Using elements of art in describing the artist's process		─	─	573	0.62**
7	In the Art Museum: Demonstrating knowledge of museum conventions		─	─	─	─
8	Art History: Demonstrating an understanding of art history chronology		─	─	─	─
9	The Artistic Process: Observing an artist at work and identifying the artistic process		─	─	─	─
10	Artistic Expression: Written response on what inspires an artist		─	─	495	0.59**
11.1	Elements of Visual Art: Looking at artwork and identifying elements in visual art	Geometric Shapes	─	─	─	─
		Organic Shapes	─	─	─	─
11.2		Negative and Positive Space	─	─	─	─
12.1	Art Making: Creating a work of art	Uses a Variety of Textures	588	0.55**	─	─
12.2		Uses a Variety of Colors	─	─	573	0.38**
12.3		Uses a Variety of Shapes	588	0.41	573	0.47**
12.4		Uses Space Appropriately	589	0.52**	574	0.53**
12.5		Clearly Demonstrates Use of Negative and Positive Space	589	0.37**	574	0.53**
12.6		Handles Materials Competently	589	0.40**	574	0.52**
12.7		Achieves Expressive Quality	589	0.42**	─	─
12.8		Demonstrates Imaginative	─	─	573	0.67**

Task Number	Task and Description		Fall 2011		Spring 2012	
			N Cases	Kappa	N Cases	Kappa
		Capacities				
12.9		Depicts Figure and Setting	588	0.45**	569	0.62**
13.1		Reflection on Realism	564	0.44**	—	—
13.2		Reflection on Use of Texture	562	0.41**	—	—
13.3		Reflection on Use of Negative Space	556	0.43**	—	—
13.4		Reflection on Use of Positive Space	556	0.41**	—	—
13.5	Reflective of Artwork: Reflection on student's art making process	Writes in complete sentences	569	0.50**	—	—
13.6		Spells Words Correctly	569	0.48**	—	—
13.7		Uses Correct Punctuation and Grammar	568	0.46**	—	—
13.8		Reflection of Artwork	—	—	564	0.64**
13.9		Visual Arts Vocabulary	—	—	563	0.70**
13.10		Writing Skills	—	—	562	0.65**

*$p < .05$, **$p < .001$.

Table A29.
*Inter-rater Reliabilities of the **2012-2013** Fifth-Grade Benchmark Visual Arts Assessments*

Task Number	Task and Description		Fall 2012		Spring 2013	
			N Cases	Kappa	N Cases	Kappa
1.1	Elements of Art ─ Drawing Lines: Demonstrating knowledge of a variety of lines		─	─	─	─
2.1	Elements of Art ─ Visual Textures: Demonstrating knowledge of visual textures	Identifying Visual Textures	─	─	─	─
2.2		Creating Visual Textures	─	─	─	─
3.1	Drawing: Using lines and textures to create a drawing reflecting imaginative capacities	Handles Materials Competently	550	0.72**	509	0.32**
3.2		Includes Detail	550	0.65**	509	0.42**
3.3		Uses a Variety of Lines	549	0.65**	509	0.25**
3.4		Uses Visual Texture	550	0.54**	509	0.27**
3.5		Fills Page Appropriately	550	0.70**	509	0.28**
3.6		Demonstrates Imaginative Capacities	549	0.66**	509	0.38**
4.1	Color Theory: Demonstrating knowledge of categories of color	Primary	─	─	─	─
4.2		Secondary	─	─	─	─
5.1	Tints and Shades: Creating tints and shades	Tints	─	─	─	─
5.2		Shades	─	─	─	─
6	Thinking Like an Artist ─ Elements of Art: Written response, Using elements of art in describing the artist's process		538	0.63**	515	0.40**
7	In the Art Museum: Demonstrating knowledge of museum conventions		─	─	─	─
8	Art History: Demonstrating an understanding of art history chronology		─	─	─	─
9	The Artistic Process: Observing an artist at work and identifying the artistic process		─	─	─	─
10	Artist's Inspiration: Written response on what inspires an artist		531	0.67**	509	0.27**
11	Elements of Visual Art: Organic and Geometric Shapes, looking at artwork and identifying organic and geometric shapes		─	─	─	─
12	Elements of Visual Art: Negative and Positive Space, looking at artwork and identifying negative and positive shapes		─	─	─	─
13.1	Art Making: Creating a work of art	Fills Page Appropriately	545	0.70**	510	0.40**
13.2		Demonstrates Use of a Variety of Colors	550	0.68**	509	0.22**
13.3		Demonstrates Use of a Variety of Shapes	548	0.63**	511	0.30**
13.4		Demonstrates Use of Negative and Positive Space	549	0.64**	511	0.23**

Task Number	Task and Description		Fall 2012		Spring 2013	
			N Cases	Kappa	N Cases	Kappa
13.5		Handles Materials Competently ─ Scissors	550	0.62**	511	0.34**
13.6		Handles Materials Competently ─ Glue	548	0.59**	511	0.32**
13.7		Demonstrates Imaginative Capacities	549	0.60**	511	0.32**
13.1	Reflective of Artwork: Reflection on student's art making process	Content	544	0.67**	506	0.29**
13.2		Comprehension	543	0.70**	506	0.25**
13.3		Usage	543	0.70**	506	0.30**

*$p < .05$, **$p < .001$.

Table A30.
*Inter-rater Reliabilities of the **2013-2014** Fifth-Grade Benchmark Visual Arts Assessments*

Task Number	Task and Description		Fall 2013		Spring 2014	
			N Cases	Kappa	N Cases	Kappa
1.1	Elements of Art ─ Drawing Lines: Demonstrating knowledge of a variety of lines		─	─	─	─
2.1	Elements of Art ─ Drawing Visual Textures: Demonstrating knowledge of visual textures	Identifying Visual Textures	─	─	─	─
2.2		Creating Visual Textures from a Masterwork	─	─	─	─
2.3		Creating Visual Textures	─	─	─	─
3.1	Drawing: Using lines and textures to create a drawing reflecting imaginative capacities	Uses Drawing Pencil with Control	546	0.27**	541	0.18**
3.2		Draws Image with High Degree of Detail	546	0.27**	541	0.33**
3.3		Uses a Variety of Lines	546	0.26**	541	0.34**
3.4		Uses Visual Texture	546	0.24**	541	0.30**
3.5		Fills Page Appropriately	545	0.26**	541	0.26**
3.6		Demonstrates Imaginative Capacities	545	0.34**	541	0.31**
3.7		Responds to Varying Demands of Audience, Task, and Purpose	─	─	541	0.46**
4.1	Color Theory: Demonstrating understanding of the color wheel	Primary	─	─	─	─
4.2		Secondary	─	─	─	─
5.1	Tints and Shades: Identifying how tints and shades are created	Tints	─	─	─	─
5.2		Shades	─	─	─	─
6.1	Thinking Like an Artist ─ Elements of Art: Written response, describing how an artist used specific elements of art to make a work of art	Use of Lines	545	0.24**	540	0.27**
6.2		Use of Color	524	0.23**	533	0.29**
7	In the Art Museum: Demonstrating ability to read object─related text		─	─	─	─
8	Art History: Demonstrating an understanding of art history. Organizing works of art chronologically		─	─	─	─
9	The Artistic Process: Examining a photograph of an artist at work and responding to a question about the artist's process		─	─	─	─
10	Artist's Inspiration: Written response relating to an artist's inspiration		545	0.27**	528	0.32**
11	Elements of Visual Art: Organic and Geometric Shapes, looking at		─	─	─	─

Task Number	Task and Description		Fall 2013		Spring 2014	
			N Cases	Kappa	N Cases	Kappa
	artwork and identifying organic and geometric shapes					
12	Elements of Visual Art: Negative and Positive Space, looking at artwork and identifying negative and positive shapes		—	—	—	—
13.1	Art Making: Creating a work of art	Fills Page Appropriately	547	0.38**	541	0.31**
13.2		Demonstrates Use of a Variety of Colors	548	0.33**	541	0.28**
13.3		Demonstrates Use of a Variety of Shapes	548	0.25**	541	0.27**
13.4		Demonstrates Use of Negative and Positive Space	548	0.27**	541	0.23**
13.5		Demonstrates Facility Using Scissors	548	0.34**	541	0.26**
13.6		Demonstrates Facility Using Glue	548	0.32**	541	0.27**
13.7		Demonstrates Imaginative Capacities	547	0.33**	541	0.35**
13.8		Responds to Varying Demands of Audience, Task, and Purpose	—	—	541	0.47**
14.1	Written Response to a Masterwork: Using evidence in a masterwork to craft a creative written response about the characters, setting, and artist's inspiration	Central Themes, Ideas, Content	519	0.36**	538	0.27**
14.2		Development	520	0.37**	538	0.31**
14.3		Analysis	520	0.38**	538	0.32**
14.4		Writing Conventions	519	0.37**	538	0.23**

Note: Tasks 1.1 through 2.3, Tasks 4.1 through 5.2, Tasks 7 through 9, and Tasks 11 and 12 do not have kappa values, given that they had right or wrong answers.
*$p < .05$, **$p < .001$.

Table A31.
*Inter-rater Reliabilities of the **2011-2012** Middle School Benchmark Visual Arts Assessments*

Task Number	Task and Description		Fall 2011		Spring 2012	
			N Cases	Kappa	N Cases	Kappa
1.1	Drawing: Drawing from observation use of shape, texture, and lighting	Uses Entire Page	297	0.61**	—	—
1.2		Renders Shape	297	0.56**	296	0.61**
1.3		Draws Detail	—	—	296	0.58**
1.4		Draws Object Large Enough to View Detail	297	0.49**	296	0.57**
1.5		Uses Visual Texture	297	0.48**	296	0.59**
1.6		Uses Light and Shadow	296	0.52**	296	0.61**
1.7		Achieve Expressive Quality	297	0.43**	295	0.62**
2	Analysis of Artwork: Comparing and contrasting two works of art		—	—	291	0.77**
3.1	Color Theory: Demonstrating knowledge of categories of color	Primary	—	—	—	—
3.2		Secondary	—	—	—	—
3.3		Complementary	—	—	—	—
3.4		Analogous	—	—	—	—
4.1	Qualities of Sculpture: Demonstrating knowledge of positive and negative space, symmetry, categories of shape, and visual arts styles	Positive and Negative Space	—	—	—	—
4.2		Symmetry	—	—	—	—
4.3		Shapes	—	—	—	—
4.4		Styles	—	—	—	—
5	Art History: Demonstrating an understanding of art history chronology		—	—	—	—
6.1	Art Making: Creating a work of art	Addresses specific art assignment	302	0.60**	299	0.68**
6.2		Uses Space Appropriately	302	0.55**	299	0.72**
6.3		Demonstrates unity through color	302	0.51**	299	0.71**
6.4		Demonstrates balances composition	302	0.55**	299	0.63**
6.5		Handles materials competently	302	0.45**	299	0.59**
6.6		Achieves expressive quality	302	0.50**	298	0.66**
7.1	Reflection of Artwork: Reflection on student's art making process	Used Elements of Art	294	0.47	—	—
7.2		Reflection of Elements of Art	293	0.39	—	—
7.4		Reflection of Artwork	—	—	289	0.66**
7.5		Expressive and Descriptive Written Response	294	0.44**	290	0.71**
7.6		Writes in Complete Sentences	294	0.50**	—	—

Task Number	Task and Description		Fall 2011		Spring 2012	
			N Cases	Kappa	N Cases	Kappa
7.7		Spells Words Correctly	294	0.57**	—	—
7.8		Uses Correct Punctuation	295	0.58**	—	—
7.9		Writing Skills	—	—	290	0.73**

*$p < .05$, **$p < .001$.

Table A32.
*Inter-rater Reliabilities of the **2012-2013** Middle School Benchmark Visual Arts Assessments*

Task Number	Task and Description		Fall 2012		Spring 2013	
			N Cases	Kappa	N Cases	Kappa
1.1	Drawing: Drawing from observation, demonstrating use of shape, texture, and light and shade; drawing details	Fills Page Appropriately	332	0.72**	359	0.46**
1.2		Uses Elements of Art to Render Shape	331	0.70**	358	0.38**
1.3		Demonstrates Use of Light and Shadow	332	0.67**	359	0.44**
1.4		Draws Details to Elicit Fully-Realized Quality	331	0.67**	359	0.43**
1.5		Demonstrates Imaginative Capacities	332	0.57**	359	0.32**
2.1	Analysis of Artwork: Comparing and contrasting two works of art	Art Vocabulary	330	0.63**	356	0.42**
2.2		Artist's Intent	330	0.64**	355	0.28**
3.1	Color Theory: Demonstrating knowledge of categories of color	Primary	─	─	─	─
3.2		Secondary	─	─	─	─
3.3		Complementary	─	─	─	─
3.4		Analogous	─	─	─	─
4	Qualities of Sculpture ─ Negative/Positive Space: Demonstrating knowledge of negative and positive space		─	─	─	─
5.1	Qualities of Sculpture ─ Symmetry, Shapes, and Visual Art Styles: Demonstrating knowledge of symmetry, shapes, and visual arts styles	Symmetry	─	─	─	─
5.2		Shapes	─	─	─	─
5.3		Visual Art Styles	─	─	─	─
6	Art History: Demonstrating an understanding of art history chronology		─	─	─	─
7.1	Art Making: Creating a work of art	Fills Page Appropriately	330	0.59**	359	0.29**
7.2		Uses Colors to Create a Balanced Composition	331	0.60**	359	0.18**
7.3		Uses Shapes to Create a Balanced Composition	331	0.56**	359	0.26**
7.4		Handles Materials Competently ─ Scissors	330	0.53**	359	0.13*
7.5		Handles Materials Competently ─ Glue	331	0.52**	359	0.14**
7.6		Achieves Expressive Quality	330	0.59**	359	0.21**
8.1	Reflection of Artwork: Reflection on student's art making process	Content	328	0.63**	352	0.24**
8.2		Comprehension	329	0.62**	353	0.32**
8.3		Usage	330	0.59**	353	0.24**

Note: Tasks 3.1 through 6 do not have kappa values, given that they had right or wrong answers.

*$p < .05$, **$p < .001$.

Table A33.
*Inter-rater Reliabilities of the **2013-2014** Middle School Benchmark Visual Arts Assessments*

Task Number	Task and Description		Fall 2013		Spring 2014	
			N Cases	Kappa	N Cases	Kappa
1.1	Drawing: Using a pencil to draw from observation, demonstrating use of shapes, texture, and light and shadow; drawing details	Fills Page Appropriately	341	0.49**	321	0.32**
1.2		Uses Elements of Art to Render Shape	341	0.35**	321	0.34**
1.3		Demonstrates Use of Light and Shadow	341	0.34**	321	0.47**
1.4		Draws Details of Object to Elicit Fully-Realized Quality	341	0.34**	321	0.43**
1.5		Demonstrates Imaginative Capacities	341	0.24**	321	0.28**
1.6		Responds to Varying Demands of Audience, Task, and Purpose	—	—	321	0.60**
2A.1	Analysis of Artwork: Looking at artwork and identifying the artist's intent, use of color, mood of artwork, and background	Artwork #1: Intent	305	0.31**	312	0.17**
2A.2		Artwork #1: Color	331	0.16**	309	0.17**
2A.3		Artwork #1: Mood	331	0.15**	311	0.15*
2A.4		Artwork #1: Background	325	0.10*	315	0.17**
2B.1		Artwork #2: Intent	335	0.23**	307	0.14**
2B.2		Artwork #2: Color	332	0.15**	301	0.20**
2B.3		Artwork #2: Mood	328	0.13*	308	0.11*
2B.4		Artwork #2: Background	321	0.13*	311	0.18**
3.1	Color Theory: Demonstrating knowledge of the color wheel	Primary	—	—	—	—
3.2		Secondary	—	—	—	—
3.3		Complementary	—	—	—	—
3.4		Analogous	—	—	—	—
4	Qualities of Sculpture ─ Negative/Positive Space: Demonstrating knowledge of negative and positive space		—	—	—	—
5.1	Qualities of Sculpture ─ Symmetry, Shapes, and Visual Art Styles: Demonstrating knowledge of symmetrical and asymmetrical balance, organic and geometric shapes, and abstract and realistic styles	Symmetrical and Asymmetrical Balance	—	—	—	—
5.2		Organic and Geometric Shapes	—	—	—	—
5.3		Abstract and Realistic Styles	—	—	—	—
6	Art History: Demonstrating an understanding of art history. Organizing works of art chronologically		—	—	—	—
7.1	Art Making: Creating a work of art	Fills Page Appropriately	339	0.40**	322	0.30**
7.2		Demonstrates Expressive Use of Colors	339	0.34**	322	0.30**

Task Number	Task and Description		Fall 2013		Spring 2014	
			N Cases	Kappa	N Cases	Kappa
7.3		Demonstrates Expressive Use of Shapes	339	0.42**	321	0.35**
7.4		Demonstrates Facility Using Scissors	339	0.28**	322	0.23**
7.5		Demonstrates Facility Using Glue	339	0.23**	322	0.35**
7.6		Responds to Varying Demands of Audience, Task, and Purpose	—	—	—	—
8.1	Written Response to a Masterwork: Using evidence in a masterwork to craft a creative written response about the characters and setting	Central Themes, Ideas, Content	319	0.28**	317	0.18**
8.2		Development	319	0.32**	317	0.24**
8.3		Analysis	319	0.31**	317	0.16**
8.4		Writing Conventions	313	0.27**	315	0.19**

*$p < .05$, **$p < .001$.

Table A34.
*Inter-rater Reliabilities of the **2011-2012** High School Benchmark Visual Arts Assessments*

Task Number	Task and Description		Fall 2011		Spring 2012	
			N Cases	Kappa	N Cases	Kappa
1	Three Dimensionality: Looking at artwork and describing artist's technique		–	–	243	0.69**
2.1	Elements of Art and Principles of Design: Writing about artwork in terms of Elements of Art and Principles of Design	Elements of Art	–	–	245	0.74**
2.2		Principles of Design	–	–	237	0.66**
3	Visual Art Vocabulary: Identifying vocabulary for art media		–	–	–	–
4	Art History: Demonstrating an understanding of art history chronology		–	–	–	–
5	Color Theory: Demonstrating knowledge of categories of color		–	–	–	–
6	Art in the Community: Written response on the relationship between artwork and its environment		–	–	231	0.71**
7.1	Curating an Art Exhibit: Choosing artwork by self-selected theme and describing choices	Art Exhibit	256	0.46**	237	0.61**
7.2		Writes for Intended Audience	–	–	225	0.64**
7.3		Expressive and Descriptive Written Response	245	0.62**	225	0.73**
7.4		Includes Art Vocabulary	242	0.57**	–	–
7.5		Writes in Complete Sentences	244	0.48**	–	–
		Spells Words Correctly	246	0.41**	–	–
7.6		Uses Correct Punctuation and Grammar	246	0.49**	–	–
7.7		Writing Skills	–	–	225	0.73**
8.1	Art Making: Creating a work of art	Addresses Specific Art Assignment	264	0.54**	240	0.67**
8.2		Uses Space Appropriately	264	0.56**	240	0.66**
8.3		Demonstrates Three-Dimensionality and Volume	257	0.53**	240	0.75**
8.4		Composition	–	–	240	0.66**
8.5		Use of Materials	–	–	239	0.70**
8.6		Demonstrates Realism	–	–	240	0.79**
8.7		Achieves Expressive Quality	261	0.57**	–	–
8.8		Demonstrates Perspective and Scale	262	0.64**	240	0.66**

*$p < .05$, **$p < .001$.

Table A35.
*Inter-rater Reliabilities of the **2012-2013** High School Benchmark Visual Arts Assessments*

Task Number	Task and Description		Fall 2012		Spring 2013	
			N Cases	Kappa	N Cases	Kappa
1	Three Dimensionality: Looking at artwork and describing artist's technique		338	0.81**	343	0.35**
2	Elements of Art: Writing about artwork in terms of Elements of Art		335	0.81**	344	0.26**
3	Principles of Design: Writing about artwork in terms of Principles of Design		325	0.74**	339	0.20**
4	Visual Art Vocabulary: Identifying vocabulary for art media		–	–	–	–
5	Art History: Demonstrating an understanding of art history chronology		–	–	–	–
6	Color Theory: Demonstrating knowledge of categories of color		–	–	–	–
7.1	Curating an Art Exhibit: Choosing artwork by self-selected theme and describing choices	Selection of Artwork	326	0.70**	328	0.22**
7.2		Content	328	0.68**	319	0.21**
7.3		Comprehension	328	0.76**	318	0.25**
7.4		Usage	328	0.68**	319	0.22**
8.1	Art Making: Creating a work of art	Addresses Specific Art Assignment	327	0.67**	336	0.22**
8.2		Demonstrates Use of Shading	328	0.81**	336	0.41**
8.3		Demonstrates Perspective and Scale	251	0.66**	336	0.48**
8.4		Demonstrates Three-Dimensionality	327	0.73**	335	0.39**
8.5		Fills Page Appropriately	328	0.69**	336	0.26**
8.6		Handles Materials Competently	328	0.71**	336	0.41**
8.7		Achieves Realistic Quality	328	0.70**	336	0.47**
8.8		Demonstrates Imaginative Capacities	327	0.69**	336	0.41**

*$p < .05$, **$p < .001$.

Table A36.
*Inter-rater Reliabilities of the **2013-2014** High School Benchmark Visual Arts Assessments*

Task Number	Task and Description		Fall 2013		Spring 2014	
			N Cases	Kappa	N Cases	Kappa
1	Three Dimensionality: Looking at artwork and describing artist's technique		320	0.42**	305	0.49**
2.1	Drawing: Using a pencil to draw from observation	Addresses Specific Art Assignment	326	0.36**	306	0.47**
2.2		Demonstrates Use of Shading	326	0.45**	306	0.56**
2.3		Demonstrates Perspective and Scale	326	0.44**	306	0.46**
2.4		Fills Page Appropriately	326	0.35**	306	0.39**
2.5		Uses Drawing Pencil with Control	325	0.36**	306	0.24**
2.6		Responds to Varying Demands of Audience, Task, and Purpose	–	–	306	0.48**
3	Elements of Art: Selecting two Elements of Art to describe how a particular design principle was achieved in a masterwork		313	0.30**	305	0.52**
4	Principles of Design: Selecting one Principle of Design to describe the composition of a masterwork		299	0.29**	296	0.37**
5	Visual Art Vocabulary: Demonstrating knowledge of art media vocabulary		–	–	–	–
6	Art History: Demonstrating an understanding of art history. Organizing works of art chronologically		–	–	–	–
7	Color Theory: Demonstrating knowledge of categories of color		–	–	–	–
8.1	Curating an Art Exhibit: Selecting related artworks from an assortment of images; justifying curatorial choice in a related essay	Central Themes, Ideas, Content	273	0.36**	288	0.47**
8.2		Development	262	0.36**	288	0.46**
8.3		Analysis	262	0.35**	288	0.44**
8.4		Writing Conventions	265	0.36**	288	0.34**

*$p < .05$, **$p < .001$.