Title:
Embracing the Burden of Proof: New Strategies for Determining Predictive Links Between Arts Integration Teacher Professional Development, Student Arts Learning, and Student Academic Achievement Outcomes

Author:
Scripp, Lawrence, New England Conservatory of Music
Paradis, Laura, Chicago Arts Partnerships in Education

Publication Date:
2014

Author Bio:
Educated as a musical performer, composer, and psychologist of music and education, Dr. Larry Scripp is the Founding Director of the Music-in-Education Concentration at New England Conservatory and the founding director and senior researcher for the Center for Music and Arts in Education (CMAIE) in New York, an organization that provides 'guided practice' consultant services for organizations that embrace arts and arts integration practices, program evaluation, and research design, data collection, and analysis.

Laura Paradis is an Independent Consultant and former Research Associate with Chicago Arts Partnerships in Education.

Abstract:
This article provides a window into Chicago Arts Partnerships in Education's (CAPE) Partnerships in Arts Integration Research (PAIR) project conducted in Chicago public schools (CPS) (pairresults.org), which statistically demonstrates how a three-year arts integration project can impact treatment versus control students in both academic and arts cluster schools. A multivariate

design framework featuring the development of survey, interview, and performance assessment instruments was used to document and rate multiple aspects of individual teacher and student performance. This design also included a series of correlation and stepwise regression analyses[i] demonstrating that statistically significant links existed between various teacher professional development outcomes, student arts and arts integration performance assessment outcomes, and academic test results. Overall, these findings offer evidence that students at schools with an arts focus combined with arts integration programming scored higher on state academic tests than did students who received exclusively academic or conventional arts learning instruction. Furthermore, these data revealed that the achievement gap between previously designated low, average, and high performing students had narrowed or disappeared. Because these findings are based on multivariate statistical methods,[ii] researchers were able to identify what sequence of factors was most predictive of achievements in student outcomes.

[i] A statistical process used to sort the single most powerful predictor of academic achievement in the context of many competing factors, which, when considered in isolation, all correlated significantly with a primary outcome variable.

[ii] Methods that allow for exploration of a broad range of possible interrelationships among variables, rather than narrow the scope of inquiry testing for simple one-way causal relationship between two variables

Embracing the Burden of Proof: New Strategies for Determining Predictive Links Between Arts Integration Teacher Professional Development, Student Arts Learning, and Student Academic Achievement Outcomes

Lawrence Scripp with contributions from Laura Yung-Chin Paradis

This article is an abridged version of the comprehensive report on the Partnerships in Arts Integration Research project conducted by Chicago Arts Partnerships in Education in the Chicago public schools[i]

Abstract

This article provides a window into Chicago Arts Partnerships in Education's (CAPE) Partnerships in Arts Integration Research (PAIR) project conducted in Chicago public schools (CPS) (pairresults.org), which statistically demonstrates how a three-year arts integration project can impact treatment versus control students in both academic and arts cluster schools. A multivariate design framework featuring the development of survey, interview, and performance assessment instruments was used to document and rate multiple aspects of individual teacher and student performance. This design also included a series of correlation and stepwise regression analyses[ii] demonstrating that statistically significant links existed between various teacher professional development outcomes, student arts and arts integration performance assessment outcomes, and academic test results. Overall, these findings offer evidence that students at schools with an arts focus combined with arts integration programming scored higher on state academic tests than did students who received exclusively academic or conventional arts learning instruction. Furthermore, these data revealed that the achievement gap between previously designated low, average, and high performing students had narrowed or disappeared. Because these findings are based on multivariate statistical methods,[iii] researchers were able to identify what sequence of factors was most predictive of achievements in student outcomes.

Introduction

Currently, arts organizations across the country and their school partners are searching for substantive evidence to confirm the impact of arts integration programs on student learning. Specifically, can arts integration professional development outcomes for individual teachers be positively linked to student arts learning? And do measures of student arts learning predict enhanced overall academic performance?

Thus far, arts learning organizations and researchers have been unable or unwilling to determine whether statistically significant causal links exist between the core elements of arts-based interventions in public schools. Often shying away from recognizing the "burden of proof" for determining the success of arts integration programs, administrators are left to describe the effectiveness of program implementation without being able to statistically connect these successes to average teacher professional development, student arts-based, or student academic-based learning outcomes.

The Need for a "Chain of Evidence" to Verify the Quality and Impact of Arts-integrated Teaching and Learning Practices in Schools

Ever since the *Champions of Change* report in 1999 (Fiske, Ed.), heads of arts organizations, public school administrators, parents, and teachers have been hearing about the positive effects of arts integration programs in public schools (Catterall, J. & Waldorf, L. 1999; Wolf, D.P., 1999). However, while evidence for the positive impact of arts integration programs has been accumulating over the past decade, research design and methods of analysis have often fallen short of establishing positive links between multiple program intervention factors and student outcomes.

Because evidence of the success of arts integration programs in public schools has been fragmented, we will continue to see headlines that suggest strong singular positive effects, such as "Test scores are raised", or "Music makes you smarter." We will also continue to encounter dismissive views claiming, "There is no definitive proof that arts cause learning in other subjects" or ambiguous statements like "What is really learned in the arts is not what you think." Yet, reports from those on the ground level of education—arts and classroom teachers, administrators and parents—continue to extol the catalytic effect of integrating arts into the curriculum. Classroom teachers assert that they have improved their teaching practices through their professional development programs. Parents and administrators use anecdotal and inferential statistical evidence to demonstrate that participation in arts programs results in higher percentages of student attendance, intrinsic motivation to learn, and more sophisticated teaching and learning practices. Additionally, neurologists demonstrate that the study of music, for example, positively affects brain development, without claiming how this information might be used to improve arts instruction. Though many are receptive to offering arts integration programs in schools based on such disparate strands of evidence, studies have not yet been designed to demonstrate definitively how connections among the various program factors can convince skeptics that high-quality arts integration programs can play an essential role in both arts and academic learning outcomes.

However, the complexity of factors involved in arts integration interventions in the public

Scripp and Paradis: Embracing the Burden of Proof: New Strategies for Determining Predicti...

3

school curriculum should not deter arts educators from searching for research methods that can detect strong associations among the factors that matter most to arts integration program directors. When teacher professional development, arts learning, and academic outcomes are not linked, program impact cannot be confirmed. For instance, reporting the percentage of teachers who are rated highly for curriculum design or implementation factors is necessary, but not sufficient, to claim that a predictive link exists between teacher quality and positive student outcomes. Without documenting and statistically measuring differences in individual teacher responses to professional development opportunities, we cannot deduce differences in quality of teaching between the control and treatment groups. Without reliable teacher outcome data, we cannot infer that the student learning outcomes have any specific connection to particular aspects of arts integration learning. However, in CAPE's PAIR project, the researchers insisted on ranking teacher outcomes on an individual basis in order to fully understand its impact on student achievement.

If linked through a determination for significant positive correlation, each factor in the chain of evidence will support the inference of the direction of causal links between each aspect of the program, with each "link" predicting results into the next. Using stepwise regression methods, the chain of evidence linking teacher professional development to student arts learning and to academic learning can be examined for the principal predictive factors.

If, however, no significant positive correlations among the various teacher professional development and student learning outcomes exist, there would be no statistical basis to argue for the positive causal impact of arts integration intervention in public schools. If there are positive correlations between teacher professional development variables and arts learning outcomes, but not academic outcomes, educators may judge arts learning to be an essential form of learning that has no statistical relationship to academic performance.

Once several factors in a sequential chain of program events can be measured, they can be structured and statistically tested as a regression equation. This stepwise procedure allows researchers to determine the relative "degree of association" among these variables in order to determine how robust the inter-relationships are with respect to a chosen outcome variable, such as academic test scores. Furthermore, stepwise regression factor analysis allows researchers to "tease out" the strongest predictive factors in the chain of evidence for the positive impact of arts integration practices on both teacher and student learning.

The results from Chicago Arts Partnerships in Education will serve as a prime example of the multivariate methodology outlined here.

The PAIR Arts Integration Design Experiment

The three-year longitudinal Partnerships in Arts Integration Research (PAIR) Arts in Education Model Development and Dissemination (AEMDD) grant awarded to Chicago Arts Partnerships in Education (CAPE) and Chicago public schools started in 2007. PAIR placed arts integration programs in six neighborhood public elementary schools, all of which had some type of regular arts instruction. What differentiated these schools was their "cluster focus," three of which had an academic orientation—math and science; reading and literature; or world languages and cultures—and three of which had a fine and performing arts focus (see Table 1). These cluster focus schools each had on staff two lead teachers whose roles were to integrate two subjects, for instance, fine and performing arts, or math and science, throughout the K-8 curriculum. The control group schools were comparable in this respect, and both control and

3

4

treatment schools had student populations with comparable standardized test scores at the beginning of this program.

Table 1: PAIR Research Design Matrix

PAIR CAPE Arts Integration Treatment Schools		PAIR Control Schools	
Math/Science Focus	Arts Focus	Math/Science Focus	Arts Focus
ELA/Writing Focus	Arts Focus	ELA/Writing Focus	Arts Focus
World Language/Cultural Studies Focus	Arts Focus	World Language/Cultural Studies Focus	Arts Focus

In schools with the arts integration programming, teaching artists representing two different artistic disciplines were asked to co-design and co-teach 10 sessions with the classroom teacher (20 sessions total). The classroom teacher was then asked to extend the curriculum in subsequent class periods.

In the first year of the project, the teaching artists worked with fourth-grade teachers; in the second year, they worked with fourth- and fifth-grade teachers; and in the third and final year, they worked with fourth-, fifth-, and sixth-grade teachers. The project design gradually expanded arts integration programs across three years, following a primary longitudinal cohort of students with an average of 110 students in each of the control and treatment cohorts.

As a result of uneven exposure to teacher professional development experience, the teaching artists who collaborated with the classroom teachers emerged as a primary quality control factor for the project. The teaching artists were given a major role in documenting their students' work and communicating with research staff about the academic and artistic concepts presented through their curriculum. Their experience in previous CAPE programs also assured the program and research staff of continuous intervention of high-quality arts-integrated work. If the teaching artists had not remained the same from year to year, the research team would have not been able to compare student learning outcomes fairly across the three years of project implementation.

In PAIR, the staff and researchers studied the effect of arts integration programs on four categories of schools:
- Level 1: Academic cluster focus (math, writing, world cultures) with conventional arts instruction;
- Level 2: Arts cluster focus with conventional academic instruction;
- Level 3: Academic cluster focus (math, writing, world cultures) with conventional arts instruction ***plus*** arts integration program;
- Level 4: Arts cluster focus with conventional academic instruction ***plus*** arts integration program;

The PAIR Arts and Arts Integration Assessment Tools
The goal of the research design framework was to provide a better way to understand, code, and analyze the data for evidence of the impact of a distinctive brand of arts integration teaching on student learning. If results from multiple sequential factors in multivariate analyses were statistically determined to be uniformly positive, organizations could then make substantial claims for the impact of their professional development program on both arts learning and

4

Scripp and Paradis: Embracing the Burden of Proof: New Strategies for Determining Predicti...

5

academic achievement.

In PAIR, the CMAIE research staff also collaborated with CAPE program staff and teaching artists to create two assessments that were central to linking teacher and student outcomes: a snapshot of arts-integrated learning (SAIL) interview, and the arts integration portfolio conference (AIPC) interview and performance assessment protocol. These assessments measured students' knowledge of fundamental concepts and processes of arts and academic learning as well as the relationship between arts and academic learning. These scores were then measured against other program outcomes, including standardized tests scores.

By the third year of the project, these tools had been field tested for reliability of administration and coding, then validated by the teaching artists and their collaborating classroom teachers after having been revised to more accurately reflect program goals. Thus, results from the third year are presented in this article with the knowledge that they had extensive trial-and-error phases, from which field researchers and research analysts discussed means of error correction.

An Arts Integration Performance Assessment Tool: The SAIL Interview Protocol

The SAIL (Snapshots of Arts Integration Learning) interview protocol was designed for this project to be independent of the program, so that both control and treatment students' arts and arts integration learning could be measured. Built on an assessment from a prior CAPE project, the questions originated from conversations between staff and experienced teaching artists, and then were edited to reflect the arts-integrated nature of PAIR. The final questions, classified by topics, are used to measure a student's knowledge of specific artistic disciplines and academic content, and the way these are related. The questions focus on: philosophy of arts and arts integration teaching and learning, artistic process, concept of a mistake, skills employed, ability to create meaning, ability to express oneself, and ability to use imagination. Staff viewed these concepts as central to any artistic discipline.

A CAPE staff member and the onsite field researcher trained the interviewers to make sure they gave each student equal opportunity to provide examples and rich description, particularly if students responded with brief answers. Scoring the interviews also required training to ensure the reliability of the results. Student responses were scored from 0 to 4, with 0 being no relevant response, to 4 being systemic understanding.

Table 2: PAIR Snapshots of Arts/Arts Integration Learning Scoring Rubric Categories and Sample Student Response Indicators

Categories of Response	Sample Anchor Statements
Level 0: **No Relevant Response** Irrelevant or indiscernible response; silence	"Nothing," "I don't remember," "I don't know," skips question; doesn't understand or respond to the question because of language problems; unintelligible mumble; or Answer is not relevant/does not address the question.
Level 1: **Single Dimensional Responses** Concrete, un-detailed response. Generic statements, singular perspective. Unspecific, unfocused, diffused. No elaboration, no detail, no personal specifics or procedural relationships. Lists undifferentiated elements.	"I liked making buildings." "We danced together."
Level 2:	

5

6

Multiple Single Dimensions Concrete connections, some occasional detail, some elaboration, or emerging specificity. Some coordination of elements, like a clearly ordered procedure. Specific personal insight.	"We danced in different ways and we had to keep the beats." "I made drawings of buildings and then we had to make the buildings with paper."
Level 3: **Coordination of Dimensions** Detailed descriptive relationships. Often provides elaborative detailed statements. Evidence of higher-order relational thinking, including elements of *inter*-personal insight and purpose, artistic aesthetic, and/or historical references.	"We had to draw buildings with the numbers on them so we could build a building with the same shape, but much bigger." "We would all dance different motions together but it had to be a fraction too. We counted the beats so the different motion had the right fraction."
Level 4: **Systemic Understanding** Substantial detail and specificity. Causal statements. Compare and contrast relationships. Critical perspective, highly complex, multiple relationships.	"When we did the dance it was really math, too. We had to count. We had to get the fractions right. It had to be right so everyone could do it together." "Our drawings are art, but they are math too, because all the numbers add together and it has to look good, too. Sometimes we didn't do the numbers right and it didn't look right and we had to fix it."

SAIL Interview Data Analysis

When comparing the control schools to the treatment schools in overall academic performance, the results favored the treatment schools from several different perspectives, as shown in Figures 1 and 2.

The left bar graph displayed in Figure 1 shows that, on average, the treatment schools scored higher in assessments of arts and arts integration learning as indicated by the SAIL interview ratings. The center bar graph reveals that arts-focused treatment school student cohorts scored higher than those at all other types of schools, suggesting that strong "arts plus arts integration programs" result in optimizing both academic performance (as previously indicated) and artistic understanding. The bar graph on the right demonstrates that arts integration practices may affect specific types of academic focus cohorts differently. Although the math and world languages schools did outperform their control school counterparts in arts assessments, the language arts schools did not. This exception, however, may have occurred simply because the control language arts school was much more likely than math or world languages schools to employ arts integration practices similar to the PAIR treatment school program.

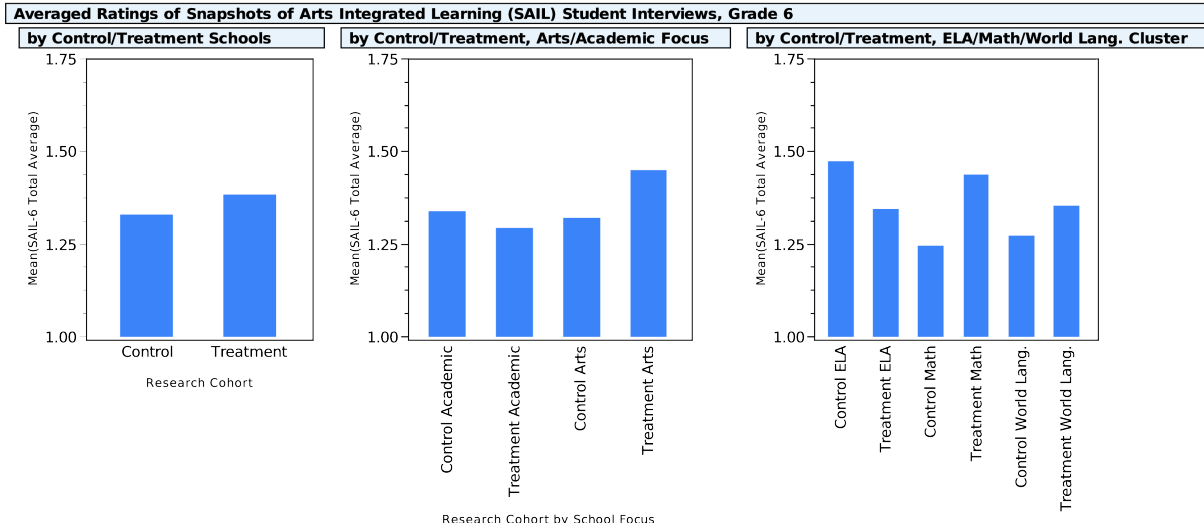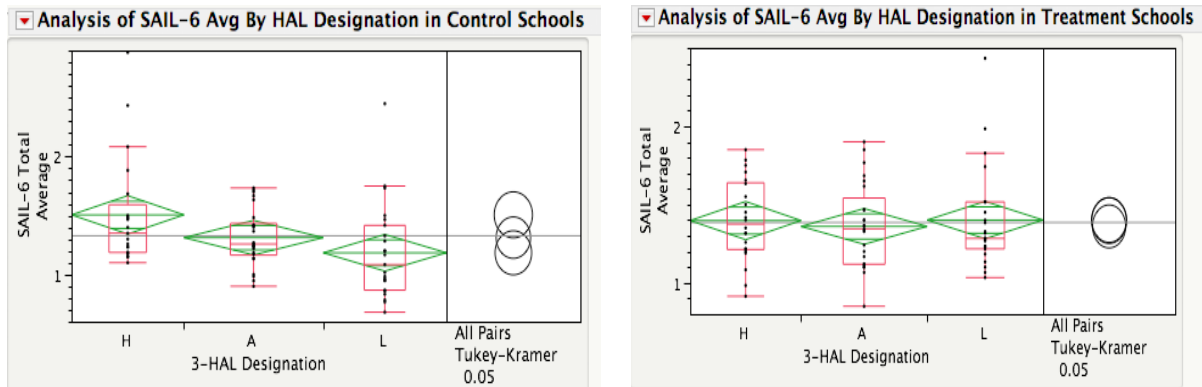Figure 1: Control-Treatment Comparisons of Grade 6 SAIL Interview

6

Scripp and Paradis: Embracing the Burden of Proof: New Strategies for Determining Predicti...

7

Figure 2 provides statistical evidence suggesting that the learning culture of the PAIR schools has been affected by CAPE's teaching artist-based arts integration program. A more equitable learning culture is characterized by a higher degree of mobility among learners, previously identified as high, average, or low academic achievers. In the treatment schools, it is evident that the pre-designated high-average-low (HAL) achieving students (based on third-grade test scores administered just before they entered the program) did *not* predict how well they would respond to the SAIL interview questions. In fact, by the final year of the project, the SAIL treatment school ratings are distributed in much the same way, regardless of their initial categorization, suggesting that positive performances in the alternative arts learning assessments were achieved by all three academically rated cohorts equally well. (See right-hand side chart of Figure 2). In the control schools, however, the arts learning scores mirrored their previous level of academic achievement. That is, the HAL student cohorts remain hierarchically ordered with respect to their third-grade ranking by the time they complete sixth grade, indicating that conventional arts education outcomes segregated from arts integration had no effect on cognitive mobility in terms of academic ranking compared to the PAIR treatment schools.

Figure 2: Control-Treatment Comparisons of SAIL Ratings by HAL Academic Designations
The Teacher-student Portfolio Conference Interview and Performance Assessment
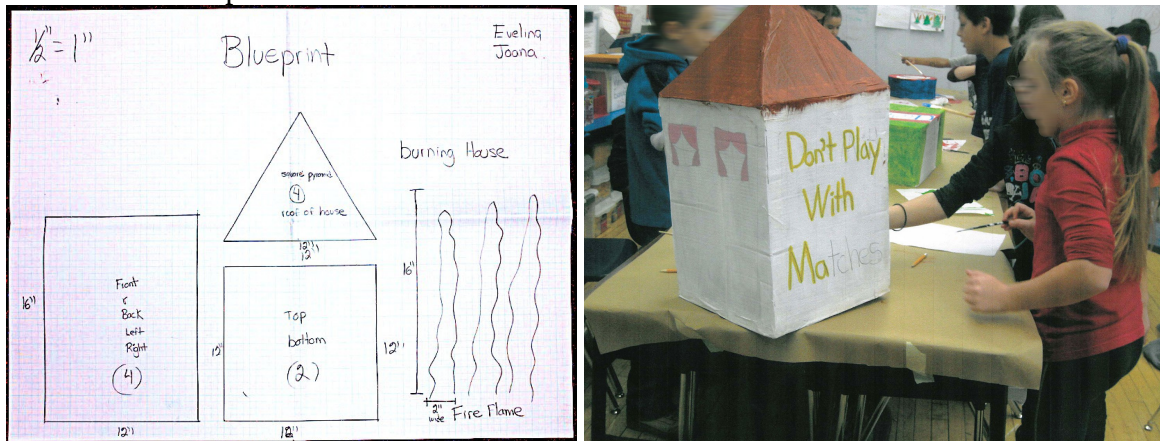

Protocol

As the central focus of PAIR teacher professional development was to find ways to document student arts integration work, teachers used documentation to share their project with other teams, to understand relationships between academic and artistic learning, and to reflect on their own teaching.

Originally conceived to evaluate teaching practices and student learning outcomes, the program and research team both discovered how the PAIR portfolio conference protocol—designed to elicit reflection on students' learning based on their portfolio work—could also serve as a tool for evaluating teachers' understanding of the teaching and learning goals of the project. As with the SAIL interview, the refinement of this portfolio conference assessment instrument was a collaborative process that required research phases, including revising the underlying research questions, changing interview protocols, and honing administration techniques and assessment rubrics before its reliability and validity could be established.

In preparation for the portfolio conference, each teacher selected three students—one from each HAL category—and made available each student's portfolio work, which could include video, group work, paper and pencil, and photographs of students performing. Neither the facilitator nor the students knew who had been previously rated as HAL students before conducting the performance assessment interviews. The questions asked were used to rate both student and teacher responses. For the first ten minutes of the conference, the facilitator asked the teacher questions about the goals of the project, and the way he or she had collaborated with the teaching artist and extended the curriculum when the teaching artist was not there. This conversation, in addition to a brief written project summary, helped the facilitator frame the dialog with students.

The middle 35 minutes of the portfolio conference focused on the students' answering questions and demonstrating aspects of their learning, with the teacher functioning as an observer charged with the task of articulating the goal of the program in the classroom and describing what the students had learned from their participation in the project. The facilitator asked these students to select at least two pieces of work—one to represent each artistic discipline. Each student was given the opportunity to talk about his or her work in each arts discipline and its connections with academic disciplines. Students were prompted also to discuss other students' work.

Figure 3: Portfolio artifacts from a unit where students built 3-dimensional sculptures from 2-dimensional blueprints

For example, in the context of a math-sculpture project (see Figure 3) students could be

Scripp and Paradis: Embracing the Burden of Proof: New Strategies for Determining Predicti...

9

asked to explain, demonstrate, and discuss the mathematical principles relevant to their sculptures. These included the design of their blueprints, the production of their three-dimensional work, and elaborations on the process or meaning of their or their peers' final work.

Students then would be asked to relate what they had learned by adopting other avenues of inquiry. For instance, students would be asked not only to demonstrate how scale was applied in their sculpture, but also to explain what the scale would be if their sculpture of a building were the size of the school, or how scale related to art making in general.

In the final minutes of the conference, students were excused from the room. The facilitator then would ask the teacher how the portfolio conference demonstrated arts integration learning, if there was anything that had surprised him/her, and if he believed an outside viewer could determine the pre-designated high, average, or low performing students. The teachers' and students' responses were transcribed and then scored by off-site researchers using the scale presented in Table 2.

In portfolio conferences, dialog with all of the participants remains the medium for assessing student responses. This dialog unfolds in a rhythm of inquiry, then initial response, and finally follow-up responses that the facilitator uses to focus on the elaboration of fundamental concepts shared between two or more disciplines. In Figure 4, the conversation about fractions evolves into a dialog that invites all students to contribute different examples of how fractions can be applied to circumstances far beyond the math text they normally use.

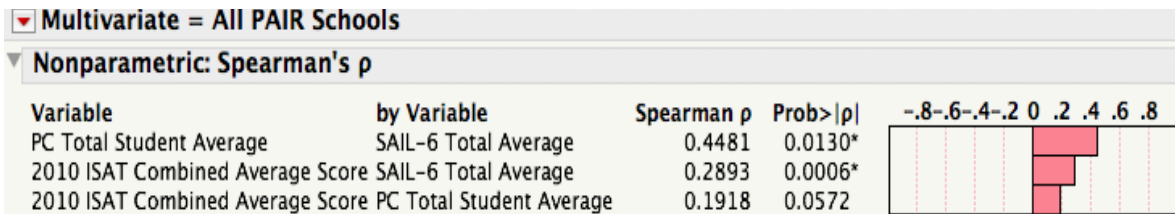Figure 4: Sample Portfolio Conference Conversation and Assessment

Student A:	I got...we had...with Miss Jessica (CAPE dance teaching artist), she … had us do two things. She made us do where we had to have a fraction of ourselves…
	First I did, like, what I am, like my race … and I'm going to say it to you like how I wrote it. "I am a fraction. I am 33% Mexican, 33% Puerto Rican, 33% Honduran, and 1% American. I am whole." She made us do this. She said we can do one part, too, and I did two. Want to hear the other one?
Facilitator:	Yes, please.
Student A:	I am a fraction. I am 50% kind, 15% mean and 35% happy. I am whole.
Facilitator:	[*Laughs.*]. I like the proportion there. Now, why did Jessica, who does dance, care about fractions in relationship to who you are? That's not even about dance. Why did she do that?
Student A:	Because for fractions, 'cause, like, how Student K was saying, we had to have different fraction counts for how much we move. And then she also kind of wanted us to do a little personality in our dance routines.
Rater Remarks:	*3.5 Connects fractions/percentages with choreographic structure and self-identity; substantially detailed explanation*

In the first segment, the conversation evolves rapidly from Student A's identity in terms of personality traits to the proportion of motions in their "dance routines" that add personality to

the choreography they created. The relatively high rating of this strand of dialog is based on Student A's ability to articulate how a mathematical concept can be used in several different ways.

Figure 5 shows that when measuring the degree of inter-correlation between the SAIL interview, the portfolio conference, and student state test scores, the SAIL and portfolio conference together are both highly associated with student academic achievement. Thus, the alternative arts learning assessments in PAIR provide both a wider spectrum of student achievement and a more coherent view of the impact of arts integration on both arts and academic assessments. The implications of these results are discussed further in the next section.

Figure 5: Degree of Correlation Between Student Portfolio Conference Ratings, Student SAIL Interview Response Ratings, and ISAT Combined Academic Test Scores

Multivariate = All PAIR Schools

Nonparametric: Spearman's ρ

Variable	by Variable	Spearman ρ	Prob>$|\rho|$	-.8-.6-.4-.2 0 .2 .4 .6 .8
PC Total Student Average	SAIL–6 Total Average	0.4481	0.0130*	
2010 ISAT Combined Average Score	SAIL–6 Total Average	0.2893	0.0006*	
2010 ISAT Combined Average Score	PC Total Student Average	0.1918	0.0572	

Putting the Puzzle Together: Mapping the Causal Connection Between Teacher Professional Development, Student Academic, and Student Arts Integration Learning Outcome Factors

Having gathered teacher professional development outcomes, student artwork, and student performance data, PAIR researchers were able to test each strand of the multivariate design framework to determine the strength of their interrelationships among the other program factors in order to determine the impact of the PAIR program on the principal outcome variables.

The SAIL interview and PAIR portfolio conference protocols were developed as the two principal alternative assessments of students' arts integration learning to be tested for their association with standardized test scores. Teacher professional development and performance outcome variables could also be tests for their predictive value for any of the student outcomes.. The variables representing seven independent forms of teacher and four forms of student learning outcomes in the PAIR program are listed in Table 3.

Table 3: PAIR Teachers and Student Variables

Teacher Professional Development Outcome Variables
A-I: TEACHER KEY EFFECT RATINGS
A-II: TEACHER YEARS OF PARTICIPATION IN PAIR
A-III: TEACHER PROFESSIONAL DEVELOPMENT SESSION ATTENDANCE
A-IV: TEACHER YEAR-END CURRICULUM AND SURVEY [YECS] RATINGS
A-V: CLASSROOM PAIR WORK SAMPLE RATINGS
A-VI: COMBINED TEACHER PROFESSIONAL DEVELOPMENT RATINGS
A-VII: TEACHER PORTFOLIO CONFERENCE INTERVIEW RESPONSE RATINGS
Student Survey Responses and Performance Outcome Variables
B-I: STUDENT SURVEY RESPONSE RATINGS
B-II: STUDENT SAIL INTERVIEW AND PERFORMANCE ASSESSMENT RATINGS

Scripp and Paradis: Embracing the Burden of Proof: New Strategies for Determining Predicti...

11

> B-III: STUDENT PORTFOLIO CONFERENCE INTERVIEW RESPONSE RATINGS
>
> B-IV: STUDENT COMBINED ISAT ACADEMIC PERFORMANCE RATINGS

Although the SAIL and PAIR portfolio conference data are the primary arts integration student learning measures in PAIR, additional data, particularly teacher performance and demographic ratings, were also key to determining the link between teacher and student outcome variable. Briefly described below are the teacher variables listed in Table 3: **Seven PAIR Teacher Outcome Variables.**

A-I: TEACHER KEY EFFECT RATINGS: The following effects were tested to show possible statistically significant correlations with student arts learning ratings or academic achievement scores:

- The content expertise effect: Teacher pedagogical content knowledge matters. Teachers who possessed more pedagogical content knowledge could better demonstrate how the academic content was relevant to the art form.

- The documenting-to-learn effect: The practice of collecting student work inspires teacher and teacher reflection.

- Teachers who regularly collect and reflect on student work are more apt to self evaluate and revise their own teaching practices

- **The fourth-grade effect**: Three years of professional development and implementation experience deepens teacher understanding of arts and arts integration teaching and learning practices.

A-II: TEACHER YEARS IN PAIR: Experience may be a highly predictive factor in PAIR student learning outcomes. As the program expanded grade coverage across a three-year period, teachers in the fourth grade had three years of professional development opportunities and implementation experience, in comparison to sixth-grade teachers, who had only one year. These data turned out be inconclusive, thus raising some doubt about differences that might pertain to years of experience with PAIR.

A-III: TEACHER ATTENDANCE: Teacher participation in professional development sessions presumably enhances the quality of program implementation. Quantitative differences in teacher level of participation in PAIR PD sessions can be examined for their effect on student learning outcomes, and, in the case proved to be more substantial than "Years of Study"

A-IV: TEACHER YEAR-END CURRICULUM SURVEY (YECS): The teacher survey questions were based on themes presented in professional development, including collaboration (either with the teaching artist or students with each other), participation in art making, and integration of the arts with academics, reflection, and dissemination. Similar to the SAIL and portfolio conferences, these surveys were tested at least once a year. Control and treatment school responses differences in these survey responses helped to differentiate the impact of "arts plus arts integration" practices in PAIR schools in comparison to the control schools. Results from teacher surveys revealed that attitudes toward collaborative curriculum design, in class teaching, and time given to student reflection on arts learning distinguished best the treatment and control group classroom culture.

A-V: CLASSROOM PAIL PAIR WORK SAMPLE RATINGS: Teachers collected student work throughout the project and also collected academic-focused work from classes that they, not the teaching artists, led. This work was rated on its quantity and quality. Due to the inconsistency of the student work provided, researchers were not able to evaluate the work on a

student learning outcome level, but rather used student work to assess teacher performance outcomes.

A-VI: COMBINED TEACHER PROFESSIONAL DEVELOPMENT RATINGS: This variable represented an average of several professional development outcome variables. It was created so that researchers could capture and rank-order differences in teacher professional development outcomes on an individual basis. The composite variable ratings represented a relatively discrete and comprehensive measure of PAIR teacher professional development outcomes, and were found to be substantially significantly predictive of arts and academic student learning outcomes.

A-VII: TEACHER PORTFOLIO CONFERENCE RATINGS: Teachers are assessed for degree of articulation of project goals, self assessment of teaching practices, and ability to describe and define evidence of arts integration teaching practices and student learning outcomes. This variable was rated so that its predictive value for arts and academic student learning could be determined statistically.

Four PAIR student learning outcome variables

B-I: STUDENT SURVEY RESPONSE RATINGS: The student questions mirrored the teacher survey questions, but were intended to track changes in treatment student perception of classroom practices in relation to the PAIR program such as collaboration, participation in art making, and integration of the arts with academics, and the presence of time for reflection on arts making and its relation to academic learning. Results from surveys in treatment did not reveal significant differences with control school student responses, yet there were indications that differences in student perception of PAIR had some degree of association with academic performance.

B-II: STUDENT SAIL INTERVIEW AND PERFORMANCE ASSESSMENT RATINGS: Students were assessed for their degree of articulation, of arts and arts integration learning practices and the differences between them. Performance assessment tasks focused on the examination of artworks required students to provide a description of art objects, the use of diverse tools and media in the art-making process, what makes something qualify as high quality art, and giving examples of what arts means to learning in schools. These quality ratings were used as a primary arts learning outcome predicted by one or more teacher outcome factors. Analysis did reveal differentiations between control and treatment school student cohorts.
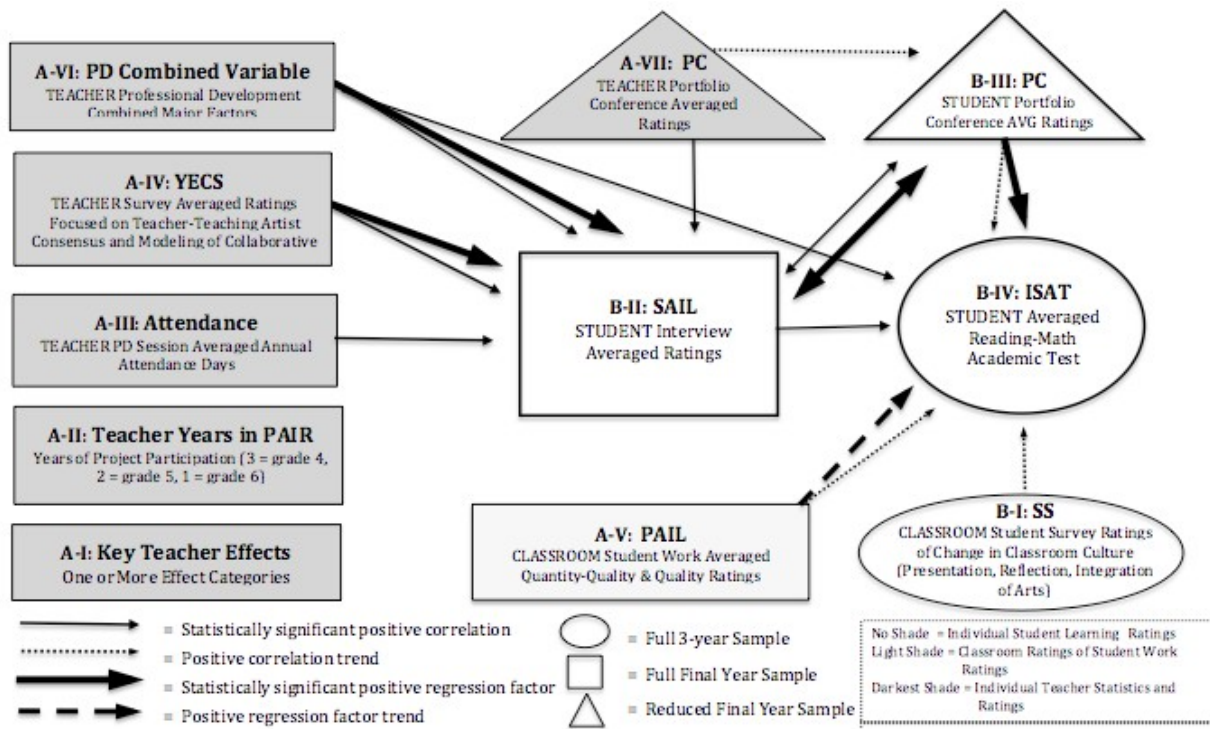
B-III: STUDENT PORTFOLIO CONFERENCE INTERVIEW RESPONSE RATINGS: Treatment school students are assessed for degree of articulation and self, peer and classroom critique of art and arts integration work, self assessment of learning practices, ability to describe what arts integrated teaching practices are, and what fellow students are learning in their arts integration classes that students in control schools may not be learning. Student scores are rank-ordered and tested for their degrees of association with academic test data and for differences between arts and arts integration assessment differences.

B-IV: STUDENT COMBINED Illinois Scholastic Achievement Tests (ISAT) ACADEMIC PERFORMANCE RATINGS: Both control and treatment School student standardized test scores are matched on baseline tests and compared for changes over the course of the three year program implementation. By year three, patterns of achievement became significantly positive for the treatment schools

Multivariate Meta-Analysis of the Statistically Significant Intersections between Teaching and Student Learning Outcome Variables by the Final Year of the PAIR Project

12

Scripp and Paradis: Embracing the Burden of Proof: New Strategies for Determining Predicti...

13

A series of correlation and regression tests performed are represented in Figure 6 to provide a perspective of the relative strengths of the program factors on student arts learning or academic learning in the PAIR project. The dotted arrows indicate significant positive statistical trends in the data; the solid lines indicate statistical significance; and the thickest lines indicate the most predictive factors in relation to all other variables, including student demographics such as gender, ethnicity, and family income.

Figure 6: A Multivariate Correlation-Regression Map of PAIR Combined Teacher-Student Outcome Intersection Factors

In the analysis of the variables depicted above, researchers concluded that the pattern and degree of correlation and regression factor analyses demonstrate that the individual teacher professional development outcome factors strongly predict student arts integration performance ratings. The teacher combined professional development factor, in particular, most highly predicts student achievement, thereby establishing causal evidence for the PAIR program's impact on student learning. The pattern and degree of correlation and, eventually, stepwise regression analyses demonstrate that the individual student arts integration assessment outcomes most directly predict academic test results, controlling for achievement, gender, family income, prior academic achievement, and ethnicity ratings (accounted for in the stepwise regression model).

When analyzing all of these factors statistically, researchers questioned how data could be looked at and quantified, and if such variables could be averaged together. Researchers paid close attention to how qualitative responses were scored to ensure consistent and valid results. In the end, we acknowledged a need for methodological experimentation in order to develop better research practices to measure teacher and student success as a result of program interventions. The findings are a result of this working approach, which will be further refined in further
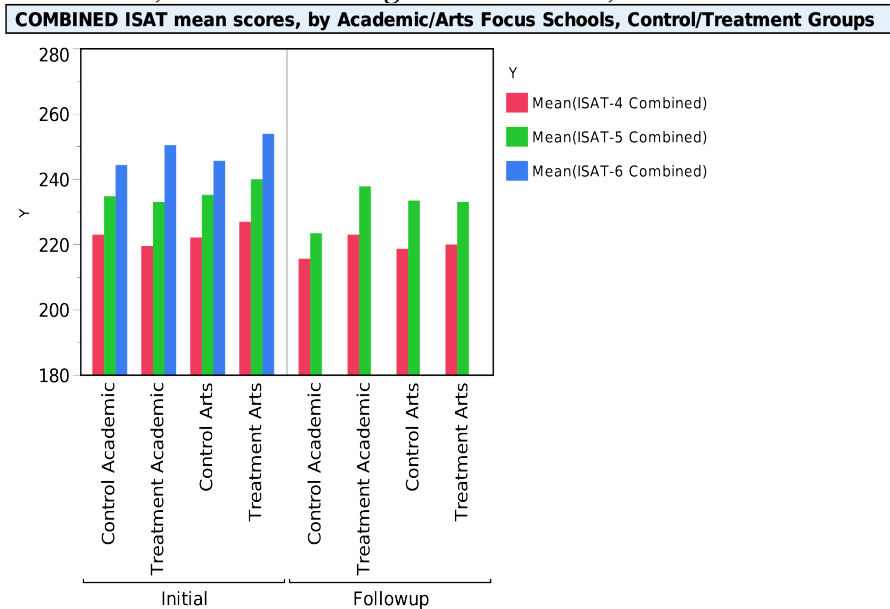
13

14

projects.

Evidence of Arts Integration's Influence on Excellence and Equity of Arts Integration Practices on Arts Learning

As demonstrated from the SAIL and portfolio conference assessments, students from treatment schools with a focus in the arts scored significantly higher on arts learning alternative assessments than did students from any other type of school. Students from treatment schools with a focus on academics frequently scored the next highest, while students from the control schools scored lower (on the SAIL). Arts plus arts integration student cohorts performed best on arts learning assessments.

In addition, students in arts integration programs in the treatment schools demonstrated greater mobility with respect to arts learning outcomes. Pre-classifications of students according to baseline third-grade HAL test scores did not predict how they performed on the SAIL by the time they were in sixth grade. Academic pre-classifications remained fixed, however, in the control schools.

The pattern of findings from the alternative arts learning assessments was also demonstrated through student academic test score comparisons (see Figure 7). That is, the longitudinal cohort in all of the treatment schools on average outperformed the control schools on averaged math-reading scores. Students from the treatment schools that combined arts integration with an arts focus were the highest rated of all of the treatment schools. This pattern was consistent for the students in the follow-up longitudinal cohort, that is, those students who participated in the program for two years instead of three. The second-tier longitudinal cohort results suggest an even greater disparity between the treatment and control school student outcomes.

Figure 7: Control-Treatment PAIR Focus School (Arts versus Academic) Comparisons of ISAT Mean Scores, PAIR Initial Longitudinal Cohorts, Grades 4-6
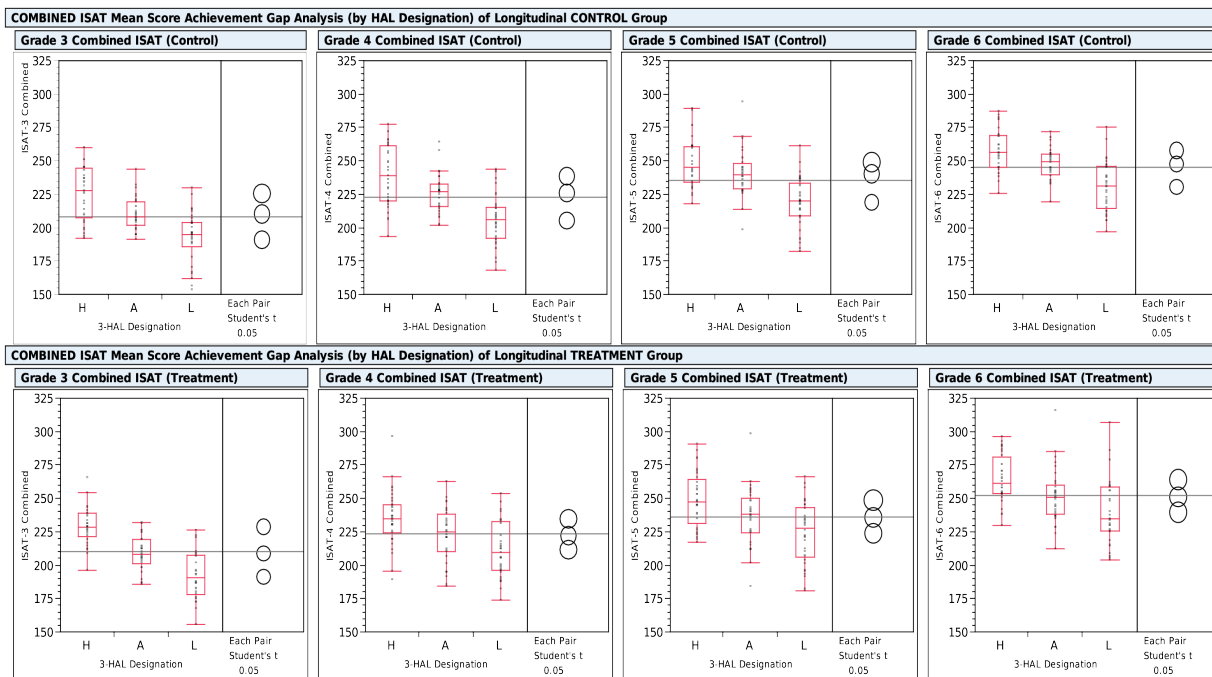

In terms of equity, low-scoring students in the control schools never closed the gap between middle- or high-scoring students, even by the sixth grade, in that there still remained a

14

Scripp and Paradis: Embracing the Burden of Proof: New Strategies for Determining Predicti...

15

statistically significant difference between the groups. Yet, in the treatment schools, the initially low-scoring students gradually approached those higher scores to the point where a significant difference in the scores no longer existed. (See Figure 8.) In fact, in the treatment school's sixth grade, several initially low-scoring arts integration students achieved scores on a par with the highest-rated control or treatment students. This phenomenon of "closing the achievement gap" in the treatment schools confirms that arts integration teaching and learning practices create greater mobility in student performance in addition to increasing academic performance.

Figure 8: Control-Treatment PAIR Schools Achievement Gap Box Plot Analysis of HAL ISAT Combined Academic Scores, Grades 3-6

Based on these performance measures, the coherency of the data, and the variables collected, researchers can begin to make recommendations for future arts-based interventions and specific changes in future research tools. As the data displays demonstrate in Figure 8, student scores do not exhibit improvements when the intervention lasts only one year. These data make a strong case for programs that give students at least two years of exposure to an arts-based intervention. Also, because student test scores are correlated with long-term, high-quality teacher participation, we also assert that arts integration practices succeed best when teachers participate consistently in professional development practices, which then lead to greater quality of student work documentation, reflective understanding of arts integration concepts, and collaboration with the same teaching artists for at least two years.

Thus, both anecdotal and statistical evidence that supports a minimum of two years of project participation is required for the program to take root and provide persuasive evidence of its impact on student learning. Multivariate analysis provides a way to understand statistically the way that high quality arts integration professional development over significant periods of time best supports arts and academic success for the student.

15

Concluding remarks: The future of research-based arts plus arts integration programs

For many years, arts learning organizations and researchers have been unprepared to determine statistically causal connections between various discrete elements of arts-based interventions in AEMDD programs. This project, however, demonstrates the results of a methodology that demands discussion and accountability for outcomes on a variety of program, teacher, and student levels. The multivariate analysis methods used in PAIR have been invaluable for studying complex learning environments in the past. The high degrees of association discussed here among a sequence of program outcomes suggests that predictive links do exist between teacher professional development and student learning outcomes in the arts and in academics.

Furthermore, this study demonstrates the reason why multiple interlocking student arts and arts integration learning assessments are necessary to determine the success of arts-based interventions in education. Taken together, they provide the basis for determining a substantive progression from program factors to teacher and student learning outcomes. In the case of the PAIR program, the SAIL and portfolio conference provided differentiated, validated measures of arts integration learning, the second of which also served to measure teacher performance and professional development.

In order for assessments to be valid and reliable, researchers and practitioners (including program staff, artists, and teachers) must work together to confirm that the assessments meet the goals of the program. Based on our analysis of these multiple outcome variables—from teacher professional development outcomes to teacher performance to student work to student learning in academics and the arts—statistical evidence now exists to support the claim that arts integration programs can improve academic performance as well as close the achievement gap. Students in the treatment schools not only outperformed students in control schools on the arts-learning performance assessments and standardized tests after three years of the program, but previously low-scoring students in arts integration schools approached the level of higher-scoring students, whereas students in the control schools remained stratified. Additionally, viewers of the PAIR portfolio conferences were often unable to distinguish between pre-classified academically high, average, and low performing students when judging these same students on the basis of their ability to demonstrate, critique, or reflect on their arts integration work. When students were offered the opportunity to demonstrate their understanding of concepts and processes shared between arts and academic learning in their arts integration projects, evidence of excellence and equity in student learning outcomes resulted. That the arts integration treatment schools outperformed both conventional arts focused and academic focused schools suggests that CAPE's arts integration practices employed in the PAIR project constitute an optimal condition for the impact of arts education on overall academic achievement.

The methodology and tools presented here lay the groundwork for arts integration practices that arts learning organizations can provide for upper elementary school grades. Future research studies at CAPE and other organizations can now build on the multivariate data collection and analyses reported here to document and assess a broad range of arts and arts integration teaching and learning practices. Together, these methods can enable school communities to embrace the burden of providing concrete evidence of the positive impact of arts integration practices on school performance, and for validating the quality and coherence of the sequential factors that exist between professional development, arts, and academic learning outcomes in schools willing to employ "arts plus arts integration" strategies for the benefit of our elementary school students.

Scripp and Paradis: Embracing the Burden of Proof: New Strategies for Determining Predicti...

17

References

Burnaford, G. (2010). Partnership in Arts Integration Research: Year Three and Final Report. Chicago, IL: Chicago Arts Partnerships in Education.

Catterall, J. & Waldorf, L. (1999). Chicago Arts Partnerships in Education: Summary Evaluation. In E. Fiske, Ed. *Champions of Change: The Impact of Arts on Learning*, 47-62.

Fiske, E., Ed. (1999). *Champions of Change: The Impact of Arts on Learning*. Washington, DC: Arts Education Partnership and the President's Committee on the Arts and the Humanities.

Scripp, L., Aprill, A., & Burnaford, G. (2007). Embracing the Challenges of Complexity in Music-in-Education Research: The Evolution of the Rubrics Cube System as a Guiding Framework for Program Development and Evaluation in MIENC schools. In L. Scripp, Ed. *Journal for Music-in-Education: Advancing Music for Changing Times*, 2, 184-201. Boston, MA: New England Conservatory.

Scripp, L. (2010). Partnership in Arts Integration Research Project: Final Reports. Chicago, IL: Chicago Arts Partnerships in Education. http://www.pairresults.org.

Wolf, D. (1999). Why the Arts Matter in Education: or Just What Do Children Learn When They Create an Opera. In E. Fiske, Ed. *Champions of Change: The Impact of Arts on Learning*, 91-98. Washington, DC: Arts Education Partnership and the President's Committee on the Arts and the Humanities.

i See pairresults.org or ArtsEdSearch.org for complete program description and analysis.

ii A statistical process used to sort the single most powerful predictor of academic achievement in the context of many competing factors, which, when considered in isolation, all correlated significantly with a primary outcome variable.

iii Methods that allow for exploration of a broad range of possible interrelationships among variables, rather than narrow the scope of inquiry testing for simple one-way causal relationship between two variables