

# Meeting the burden of proof: Statistical evidence for the impact of arts integration based on causal links among teacher professional development, arts learning and academic outcomes.

Lawrence Scripp and Laura Yung-Chin Paradis

*This paper is an abridged version of the complete 350-page technical report on the Partnerships in Arts Integration Research (PAIR) project conducted by CAPE in the Chicago Public Schools and evaluated by Gail Burnaford (Florida Atlantic University) and Lawrence Scripp (Center for Music and Arts in Education), co-Principal Investigators*

## Abstract

Having administered five different AEMDD projects in the Chicago Public Schools, CAPE has emerged as a leading proponent of multiple arts in education practices that provide professional development for both teachers and artists as they collaborate on teaching math, ELA, and social studies. CAPE's Partnership in Arts Integration Research (PAIR) grant reported in a comprehensive technical report and in this abridged article shows how a three-year arts integration intervention impacted students in 'paired' academic and arts cluster schools. PAIR researchers structured this project to account for the complexity of factors involved in student and teacher learning. The 'Rubrics Cube' multivariate research design framework was used to establish statistical links between professional development content, teacher professional development outcomes, and student arts integration and academic learning outcomes. Full completion of the research framework necessitated developing and validating additional arts integration student assessment instruments that provided evidence of learning that could be linked with teacher professional development tools and standardized academic performance ratings.

The findings from this report<sup>1</sup> demonstrate that students at schools with an arts focus combined with arts integration programming scored higher than other types of student cohorts on both state reading and math scores and other academic and arts integration learning measures. Furthermore, students in treatment schools could no longer be categorized by their initial high, average, low state score designations; the achievement gap between those students had narrowed or disappeared. Because these findings are based on a multivariable research framework, researchers can also identify the successes of professional development that led to student achievement. The results support these kinds of long term, arts integration programs — even at schools with existing art programs — as well as validate multivariable research methodologies.

## Introduction: *The need to provide a 'chain of evidence' for the quality and impact arts integrated teaching and learning*

Ever since the Champions of Change Report in 1999, heads of arts organizations, public school administrators, parents, arts and classroom teachers have been hearing about the positive effects of the arts and arts integration in education<sup>2</sup>. Projects in Chicago Public Schools (CPS) and other major cities continue to spread good news about the effects of arts and arts integration in our schools based on reports from dozens of four-year Arts in Education Model Dissemination (AEMDD) federal grants. Yet, although accumulating

---

<sup>1</sup> See the Complete PAIR report for detailed reporting of all aspects of the project ([pairresults.org](http://pairresults.org))

<sup>2</sup> See Critical Links Report, Arts Education Partnerships ([aep-arts.org](http://aep-arts.org))

evidence for the positive impact of arts integration programs has been impressive, research design and methods of analysis have often fallen short of establishing causal links between arts integration program intervention<sup>3</sup> factors and student outcomes.

When teacher professional development, arts learning and academic outcomes are not linked, studies and arts programs have limited the credibility and impact. For instance, stating the percentage of teachers with examples of the high quality curriculum is necessary, but not sufficient to claim that a link exists between teacher quality and positive student outcomes. Without documenting and statistically measuring *individual differences* in teacher response to professional development opportunities, we cannot assess differences in quality of teaching between the control and treatment groups. Without reliable teacher outcome data, we cannot assume that the student learning outcomes have any particular connection to the particular aspects of the program intervention. In PAIR, the researchers insisted on finding ways to rank order teacher outcomes on an individual basis in order to understand the impact of the professional development aspects of the program on student achievement.

Similarly, there needs to be a link between teacher professional development outcomes and student arts and arts integration learning results. Intermediary factors — measures of arts skills and understanding that directly results from the program — are key to determining the relative degree and pattern of association between teacher and student learning outcomes. This means that reliable and validated tools to measure arts learning from programs need to be developed and tested. Only after such factors are gathered can we start the process to demonstrate that the intervention program can be statistically linked with impacted student academic achievement.

Thus, the chain of evidence used to investigate the impact of arts integration interventions in schools can be understood in terms of the following hypothetical sequence of statistically significant causal links:

Multiple individual teacher professional development outcome ratings linked to—>  
Multiple individual assessment ratings of student arts/arts integration learning also can be linked also to—>  
Externally developed individual student academic outcomes (test scores).

If linked statistically, each factor in the chain of evidence sequence substantiates the direction of causal relationships, with each “link” predicting results into the next. Although the chain is linear in its explanation, the factors that construct each link do not take place in a strictly sequential fashion. The methodology and the practice itself are not entirely linear as the qualitative assessment of the teacher or student reflection on their work suggests in this study. Learning by teachers and students takes place on a continual and reiterative basis. A high statistical association between student learning outcomes

---

<sup>3</sup> Interventions imply that arts in education program represent a type of intervention in academic education meant to enhance reading or math test scores, rather than serve as enrichment programs.

and professional development outcomes, for example, reflects the impact of interactions between teachers and artists over time, and not necessarily evidence of a 'one-way' cause and effect relationship between instruction and learning. If, however, there is no significant positive correlation among the various teacher professional development and student learning outcomes, there would be no statistical basis to argue for the positive causal impact of this arts integration intervention in public schools.

Once several factors in a sequential chain of validated program outcome factors can be measured, they can consequently be structured and statistically tested as a stepwise regression equation. This statistical procedure allows researchers to determine the relative "degree of association" among these variables in order to determine how robust the causal relationships are. Furthermore, step-wise regression factor analysis allows researchers to tease out the very strongest causal factors in the chain of evidence for the positive impact of arts integration on student learning, as will be demonstrated later in this article.

#### *The 'Rubrics Cube': A Design Framework for Embracing the Complexities of Arts Integration Research<sup>4</sup>*

A significant challenge faced by PAIR researchers was finding method through which to discern relationships between the many types of outcome variables reflective of the program and distilled from the research. Not only did researchers seek to identify significant correlations between, for example, specific teacher self-reported survey questions and teacher documentation of student work, they sought to create that chain of evidence between professional development, teacher outcomes and student outcomes. The 'Rubrics Cube' research design framework, created by the Center for Music and Arts in Education (CMIAE), provided a way to understand, code and analyze the data for evidence for the impact of arts integration teaching on student learning.

As depicted below, a 'Rubrics Cube' framework depicts the three major strands of the research design and evaluation factors:

Strand 1 (top of the cube): The Rubrics Cube challenges the collaborating institutions to account for multiple forms of documentation and measurements of program outcomes as part of whole system of inter-dependent variables. The rating of the capacity of research organizations, arts learning organizations and school community partnerships to enact an action research methods (strand 3 horizontal axis below) designed to document and assess multiple program outcomes (strand 2 vertical axis).

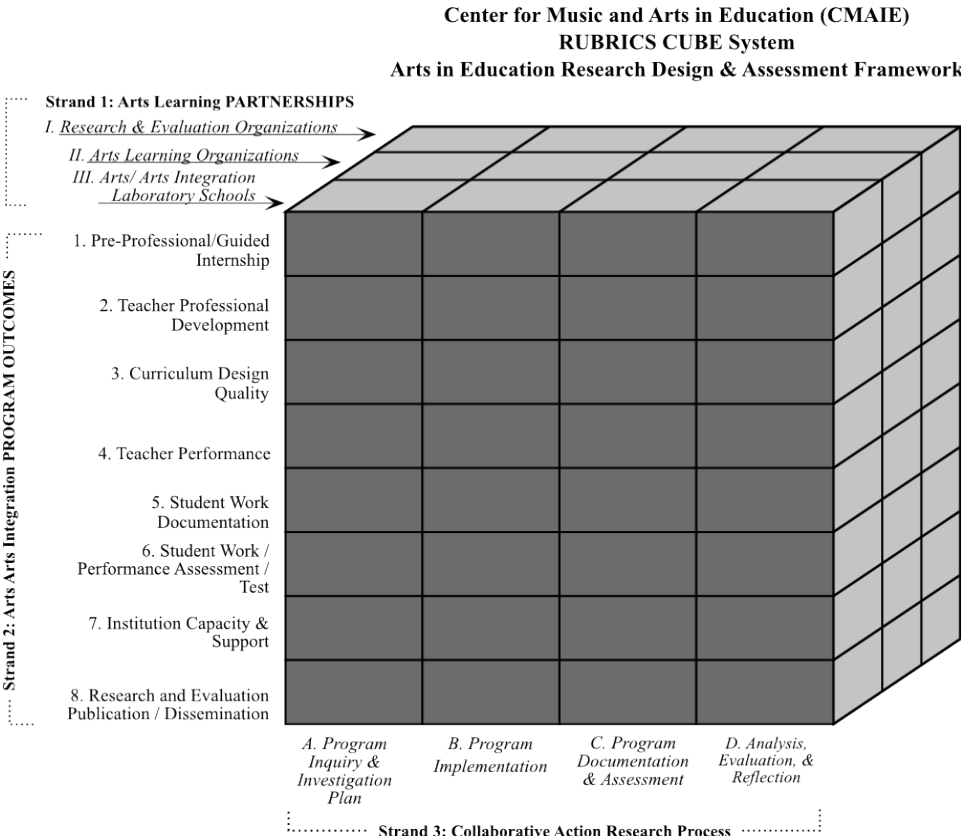
Strand 2 (vertical axis): A linear, progression of outcome multivariate factors that all play roles in the quality of program design. These factors include teacher quality based on profession development, curriculum design, curriculum implementation,

---

<sup>4</sup> Lawrence Scripp, Arnold April, et. al in the Journal for Music-in-Education (2007)

and the quality of documentation of student work outcomes—all of which play a role in the determination of measurable arts and academic student learning outcomes. Furthermore, the sustainability of the program can be predicted by the quality and effectiveness of institutional support for and capacity to enact these programs over time. The quality and timeliness of publication and reporting outcomes are critical to the dissemination and accessibility of program outcomes.

Strand 3 (horizontal axis below): The quality of the program implementation process critical to the success of the program relies on rigorous attention the documentation and evaluation of four action research phases: A. The formulation of guiding inquiry questions and plans for investigation these questions, B. Adherence to quality control checkpoints throughout the program implementation process, C. Assiduous attention to the quality of teacher and student work documentation and assessment tools, and D. Expansive analysis, evaluation, and reflection on all data collected in order to review or revise the ongoing development of the program.



In the past, AEMDD program reports have demonstrated that test scores have gone up as a result of their arts integration learning programs without providing data that links averaged individual teacher professional development outcomes with averaged individual student learning results. Most AEMDD studies have employed comparison groups or conducted control group studies to suggest that these results are not due to chance. However, studies cannot claim that a significant relationships exists between

program design and learning outcomes without establishing a chain of predictive factors that link teaching outcomes to arts learning and, by extension, to academic outcomes. Without accounting for multiple program outcomes as specified by the Rubrics Cube framework, they will not be able to (a) specify which professional development factors contributed to the quality of the program as a treatment compared to the control schools; (b) to determine to what extent student arts learning factors can be connected to professional development factors, (c) to establish the degree of correlation among multiple factors throughout a the path of program impact; and therefore (d) establish any specific basis for concluding that a particular arts integration program impacted academic achievement positively.

Thus, for the most part, evidence for the success of arts integration programs in public schools has been fragmented. Arts and classroom teachers have claimed persuasively they have developed more sophisticated teaching practices through their professional development programs. Parents and administrators use anecdotal and inferential statistical evidence to claim that participation in arts programs improves student performance. Neurologists give us evidence that the study of music positively effects brain development without claiming to know how this information might be used to improve arts instruction. Though many are receptive to offering arts and arts integration programs in schools, studies in schools have not been yet been designed to investigate links between the many variables that are necessary to convince skeptics that arts learning plays an essential role in school improvement.

If results from multiple sequential factors in multivariate analysis guided by the 'Rubrics Cube' design factors are uniformly positive, organizations then can make substantial claims for the contribution of arts and arts integration programs in our schools.

The Partnership in Arts Integration Research Project (PAIR) begins to address the issues of evaluating program outcome factors through collaborative research practice, new tool development, and multivariable analysis. In PAIR, the Center for Music and Arts in Education (CMAIE) research staff collaborated with CAPE program staff and teaching artists to create, validate, and implement additional, reliable arts learning assessments whose scores were then measured against other program outcomes, including state standardized tests scores. These additional assessments measured students' knowledge of fundamental concepts and process of arts and academic learning as well as the relationship between arts and academic learning. Students were also assessed for their ability to articulate how their work represented an integration of knowledge from multiple disciplines. Data from these assessments allowed researchers to test the methodology that emerged from dialogue between researchers and program staff, teachers and artists.

As described through the 'Rubrics Cube' framework, PAIR linked program to teacher and student outcomes statistically, and elaborated on patterns of association among various measures of teacher professional development and student results. By accounting for a wider range of variables, the researchers were able to determine indicators of both equity and excellence in student learning as a result of this arts integration program,

while controlling for students' demographic factors such as gender, race, socioeconomic status and ELL status. By determining the most potent predictors of student learning, through correlation and regression factor analysis, we can make a case for a chain of causal factors that extend from professional development condition to both arts and academic learning outcomes. PAIR demonstrated a working, collaborative model of the 'Rubrics Cube' methodology that accounts for multiple factors influencing students' learning. It identified correlating factors, measured their degree of correlation and established a statistical basis through which we can better study and communicate the impact of these arts-based interventions.

### *The PAIR Design Experiment*

The four-year longitudinal Partnership in Arts Integration Research (PAIR) Arts in Education Model 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 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 focus — Math and Science; Reading and Literature; or World Languages and Cultures — and three of which had a Fine and Performing Arts focus. These "cluster focus" schools each staffed two lead teacher positions whose roles were to integrate, 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 treatment schools had student populations with comparable test scores at the beginning of this program.

PAIR Research Design Table

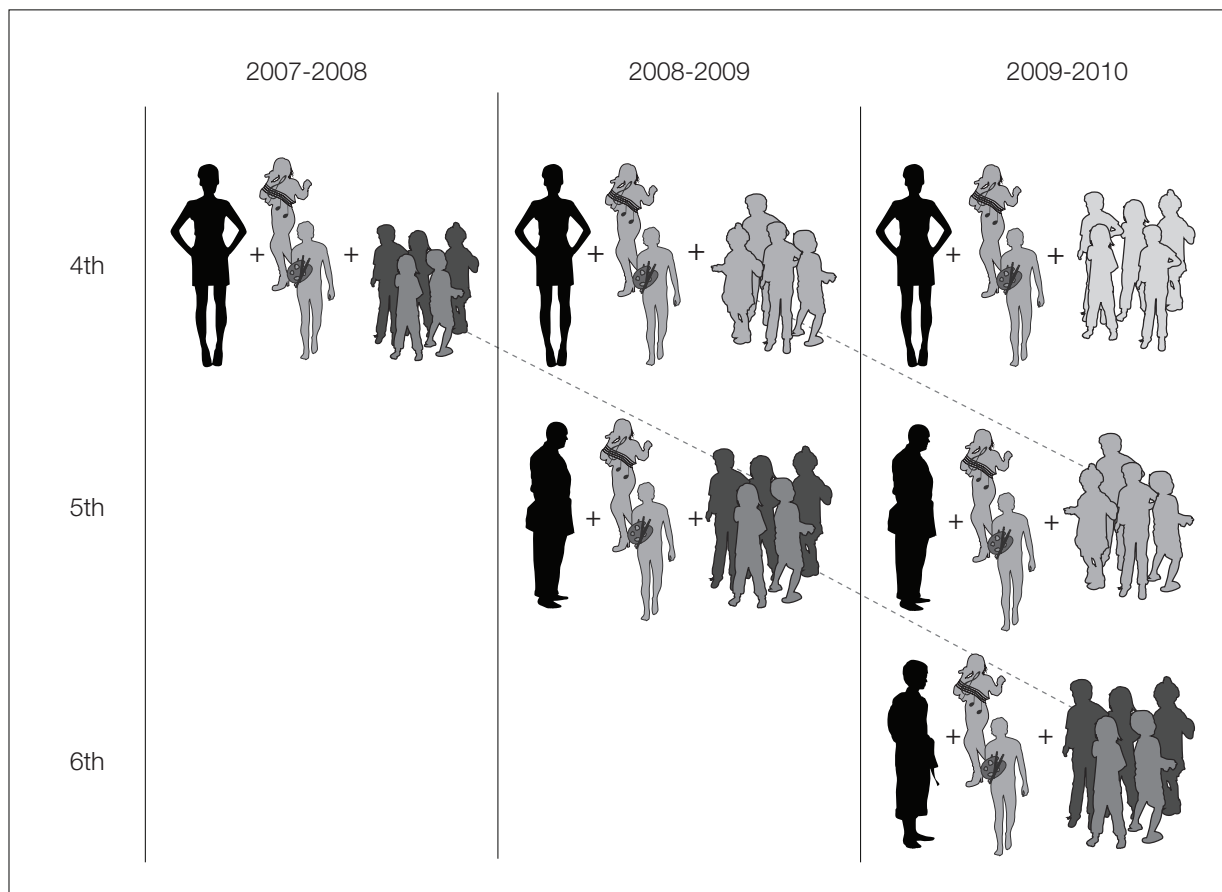
PAIR CAPE Arts Integration Treatment Schools		PAIR Control Schools	
Math/Science Focus	Arts Focus	Math/Science Focus	Arts Focus
ELA/Writing Focus	Arts Focus	ELA/Writing Focus	Arts Focus
World Language/Cultural Studies Focus	Arts Focus	World Language/Cultural Studies Focus	Arts Focus

In schools with the arts integration programming, two teaching artists representing two different artistic disciplines were asked to each co-design and co-teach 10 sessions with the classroom teacher (20 sessions total). The classroom teacher was then asked to extend the curriculum, when the teaching artist was not present as part of the co-teaching pedagogy.

In the first year of the project, the teaching artists worked with 4<sup>th</sup> grade teachers;; in the second year, they worked with 4<sup>th</sup> and 5<sup>th</sup> grade teachers; in the third and final year, they worked with 4<sup>th</sup>, 5<sup>th</sup>, and 6<sup>th</sup> grade teachers. The project design gradually expanded arts integration programs across three years, following a primary longitudinal cohort of

students from the 4<sup>th</sup>, to 5<sup>th</sup>, to 6<sup>th</sup> grade (on average 140 students in each of the control and treatment cohorts over the three years of the project). Consequently, the 4<sup>th</sup> grade teachers ended with three years of professional development and implementation experience, while the 6<sup>th</sup> grade teachers ended the program with only one. Thus, the longitudinal cohort of students who received the most arts integration instruction were also the students who were never taught by an experienced classroom PAIR teacher. This design element flaw suggests that the positive comparisons between the treatment and treatment and control longitudinal student cohorts may have been even more impressive had all classroom treatment school teachers received PAIR professional development from the onset of the project<sup>5</sup>. In this case the constant presence of the ever more experienced teaching artists ensured that the 6<sup>th</sup> grade students benefitted from highly experienced CAPE teaching artists.

Treatment School Longitudinal Design  
*Each classroom, the same two artists, adding classrooms over three years*



*Thus the primary and constant quality control factor for the longitudinal research design was the teaching artist. Despite the increase in teaching sessions, the teaching artists*

<sup>5</sup> This design flaw has since been rectified in follow-up CAPE studies now in progress.

remained the same over three years. They were given a major role in documenting their students' work and communicating with research staff 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, and if they had visible differences in their levels of experience, the research team would have not been able to compare student learning outcomes fairly across the three years of the study.

This is not a study comparing schools with arts against schools without the arts. In CPS, the average district school academic scores are not comparable with the fine and performing arts focused or academic focused cluster schools. The average academic score ratings for the district schools fall far below the average performance of all PAIR treatment and control schools in this study. In CPS, both schools that use the arts as a primary focus and schools that use selective academic focus combined with arts instruction demonstrate better school performance when compared to schools that neither have arts nor arts integration programs. In PAIR, the staff and researchers studied the effect of arts integration programs on four categories of schools:

PAIR levels of Academic and Arts/Arts Integration Programming, represented by four types of schools involved in the PAIR research and evaluation project, all included existing conventional art instruction:

Level 4: Arts cluster focus with conventional academic instruction <b>plus</b> arts integration program; Level 3: Academic cluster focus (Math, Writing, World Cultures) with conventional arts instruction <b>plus</b> arts integration program; Level 2: Arts cluster focus with conventional academic instruction; Level 1: Academic cluster focus (Math, Writing, World Cultures) with conventional arts instruction.
--

### *Collaborative Research Methods*

The PAIR research-based program design emphasizes a collaborative relationship between researchers and practitioners (program staff, teachers and teaching artists) who organize and implement these arts integration programs. A member of CAPE staff conducted the field research and served as liaisons with Principal Investigators, program staff, teaching artists, and teachers regarding research goals. This consistent communication across multiple channels (people, office and schools) established an ongoing dialogue to strengthen organizational research capacity.

From the beginning, the research team adapted different and validated measures of teacher and student learning. There were distinct tools for professional development outcomes, for teacher program outcomes, and for student outcomes. Although the tools themselves were each based on valid measures important to the field, by themselves



these instruments were not sufficient to represent the complexity nor test the coherency of the project as the whole. Although the tools eventually demonstrated indications of “success,” there was no initial attempt to measure the relationship from one tool to another.

Thus tools were continually refined year after year as part of an action research process, one that benefitted from both formative and summative analysis of teacher and teaching artist collaboration and its effect on student learning data. Instead of settling only for measures of success as identified by preexisting standards and previous hypotheses, the research team looked closely at the program’s professional development themes: collaboration, reflection, documentation, and dissemination of work. Those themes then determined the basis for revised surveys that described teacher and classroom outcomes. The new language used paid particular attention to the variety of means that teachers and artists might practice these themes, since each teacher-artist collaboration, each classroom, and each school represents a different ecosystem for the design and delivery of arts integration practices. In the case of the PAIR data analyses, averaging teacher professional development outcomes among survey responses, curriculum design, or documentation and interpretation of student learning proved to be a more sensitive, reliable and valid predictor of artistic and academic success.

Once all these data were collected and analyzed separately, Dr. Scripp spent another year with his research team to develop multivariate analytic methods for studying the interactions of teacher and student outcome ratings. Additional and recoded variables provided a series of tests of intended effects of the program and the overall sequence of causal elements that extended from multiple teacher professional development outcomes to various forms of student arts, academic and arts integrated learning results.

### *The PAIR Arts Integration Assessment Tools*

Two additional assessments were created, validated, and central to linking teacher and student outcomes: (1) Snapshot of Arts Integrated Learning (SAIL) interview and (2) the Arts Integration Portfolio Conference (AIPC) interview and performance assessment protocol. When used together they provide a coherent measure of teacher and student performance of understanding of arts and arts integration learning processes.

These internally developed assessments were developed with the local project in mind and underwent significant testing phases. In the first year of the project, data was unusable from the SAIL interviews due to lack of interviewer inter-reliability and wording of protocols. The first year of administering the portfolio conference also led to unusable data for analysis. By the third year, however, the tools had been edited in their language, as well as edited to more accurately reflect program goals. 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 correcting for error. It was only through practice and dialogue about the tools the program staff and research team were able to create validated assessments appropriate for this project.

## The SAIL Interview Protocol

Snapshot of Arts Integrated Learning (SAIL) 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 a former assessment from a prior CAPE project, the questions were originally designed from conversations between program staff and experienced teaching artists, and then 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 how those are related. The topics are: philosophy, process, concept of a mistake, skills employed, ability to create meaning, ability to express yourself, ability to use imagination. Staff concluded these topics as central concepts 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 reliability of the results. Student responses were scored on a scale from 0-4, with 0 being no relevant response to 4 being systemic understanding.

### Abridged SAIL Scoring System

Categories of Response	Sample Anchor Statements
<b>Level 0:</b> <b>No Relevant Response</b>  Irrelevant or indiscernible response; silence;	Answers, "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.
<b>Level 1:</b> <b>Single Dimensional Responses</b>  Concrete, un-detailed response. Generic statements, singular perspective. Unspecific, unfocused, diffused. No elaboration, no detail, no personal specifics or procedural relationships. Lists undifferentiated elements. "One dimensional thinking"	'We made up stories' 'I liked making up songs' 'We danced together/ 'I drew pictures of my family' 'We played drums with Charlie. It was fun.'
<b>i.Level 2:</b> <b>ii.Multiple Single Dimensions</b>  Concrete connections, some occasional detail, some elaboration, or emerging specificity; Some coordination of elements, like a clearly ordered procedure. Specific personal insight.	'We acted out stories from books, then we acted out our own stories.' 'We made up songs for our own country and sang them with bells.' 'I drew picture of food I like. I liked different fruits than other kids.' 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' 'We made up words and then we made up beats to go with them'
<b>Level 3:</b> <b>Coordination of Dimensions</b>  Detailed descriptive relationships. Often provides elaborative detailed statements. Evidence of higher-order relational thinking, including elements of <i>inter</i> -personal insight and purpose, artistic aesthetic, and/or historical references.	'We had to make draw buildings with the numbers on them so we could build a building with the same shape, but much bigger' 'Our song expressed the feeling of our new country, so we all had to like the notes and the beats of the song and sing it together' 'We made up our own beats for the characters in the story, and then the beats would change if some got mad or sad.' 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.
<b>Level 4:</b> <b>Systemic Understanding</b>	'Story telling is better when we acted them out cause you can see how all the characters move and talk and argue with each other. My story got better because we had fun making the story funnier when we did it for

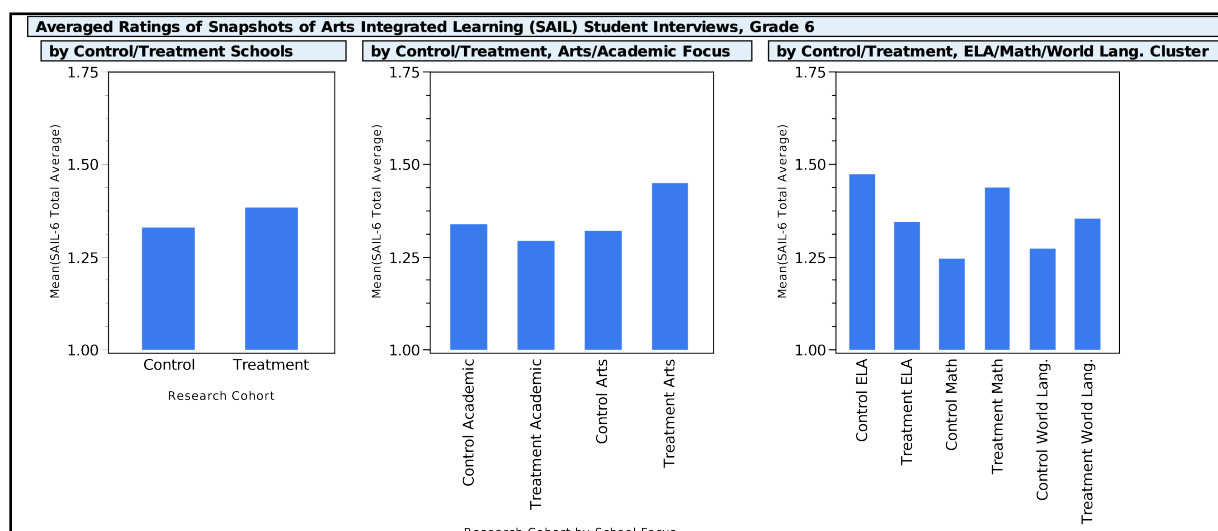
Substantial detail and specificity. Causal statements. Compare and contrast relationships. Critical perspective, highly complex, multiple relationships.	<p>the class.'</p> <p>'When we did the dance it was really math, too. We had to count. We had to get the fractions right. It had to be right so everyone could do it together'</p> <p>'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.</p> <p>"Our song expressed the feelings of the words and told about what our country is. The most important words got the highest notes so everyone would know what our country stands for. And it wasn't done until we could all sing together and that was hard to do"</p>
--	---

### SAIL Data Analysis

When comparing the control to the treatment schools in overall academic performance, the results favored the treatment schools from several different perspectives as shown in the two figures below.

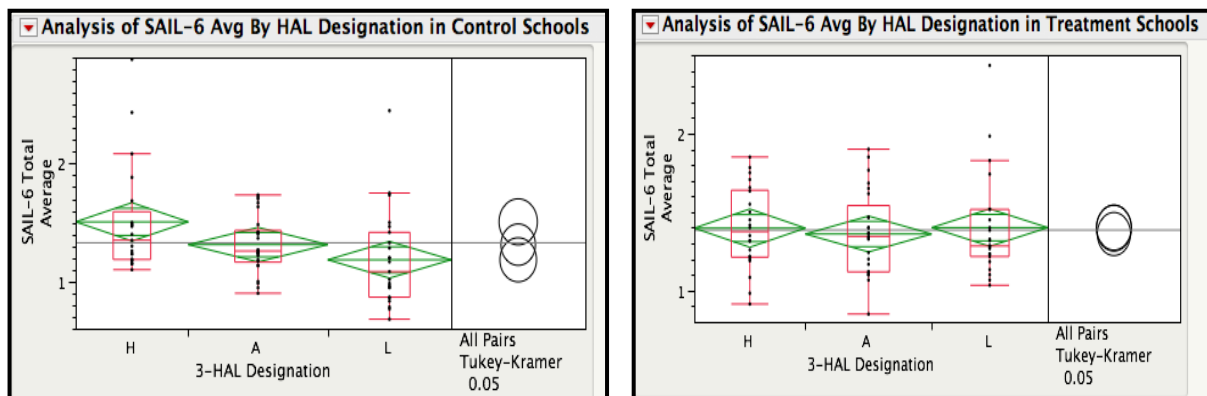
The first data display below on the left shows that, on the average, the treatment schools scored higher in assessments of arts and arts integration learning as indicated by the SAIL interview ratings. The center data display reveals that arts focused treatment school student cohorts scored higher than 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 chart on the right shows that arts integration practices may affect specific types of academic focus cohorts differently. Although the math and world language/culture treatment 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 Language Schools to employ arts integration practices similar to the PAIR treatment school program.

### Control-Treatment Comparisons of Grade 6 SAIL Interview (Final Report 3A-1 Figure 1)



The data in the figure below provide statistical evidence suggesting that the culture of the PAIR schools has been affected by the teaching artist-based arts integration program. In the treatment schools, the pre-designated High-Average-Low (HAL) students (based on 3<sup>rd</sup> 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 the arts learning reached all three academically rating cohorts equally well (see right chart below). In the control schools, however, the High-Average-Low student cohorts remain hierarchically ordered according to their third grade HAL scores by the time they leave 6<sup>th</sup> grade—a result indicating that arts learning simply reflected another measure of academic skills in the non arts integration schools and not evidence of learning in an entirely different domain of study that the PAIR schools displayed.

### **Control-Treatment Comparisons of SAIL Ratings by HAL Academic Designations** (Final report 3A-3 Figure 1)



### *The Portfolio Conference Interview and Performance Assessment Protocol*

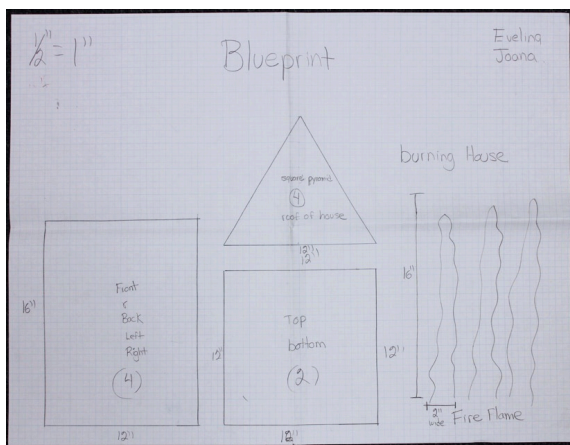
A central focus of PAIR teacher professional development was documentation of student arts integration work. PAIR teachers practiced recording, looking at, and reflecting on student work — both artistic work as well as academic work. They used documentation to share their project with other teams, to understand relationships between academic and artistic learning, and to reflect on their own teaching. Each year, they were asked to collect their students' academic and artwork as portfolios in plastic bins, organized by beginning, middle and end of project.

Originally conceived to evaluate teaching practices and student learning outcomes, the program and research team both discovered how the conference protocol — a discussion and demonstration of students learning based on their portfolio work — could also serve as transformative professional development experience for teachers. If teachers could see what their students learned using their work as a prompt, teachers might also recognize the role that documentation of arts integrated learning can play in the

classroom. As with the SAIL interview, the refinement of this portfolio conference assessment instrument was a collaborative process between researchers and practitioners that required action research phases. These included revision of the underlining research questions, changing interview protocols, honing administration techniques and assessment rubrics.

In preparation for the portfolio conference protocol, each teacher selected three students — one high, one average, one low test score performing student — and made available each students' portfolio work— which could include video, group work, paper and pencil and photographs of students performing. These students did not know they had been rated as HAL students, nor did the facilitator know how these students were classified. 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, how he/she collaborated with the teaching artist and extended the curriculum when the teaching artist was not there. This conversation, in addition to a written brief project summary, helped the facilitator frame the dialogue with students in a more expedient fashion.

The middle 35 minutes of the protocol focused on the students answering questions and demonstrating aspects of their learning, with the teacher functioning only as an observer. The facilitator asked these students to select at least two pieces of work — one to represent each artistic discipline. Each student was given opportunities to talk about his/her work in each arts discipline and its connections with academic disciplines. Students were prompted also to discuss other students' work, and to reflect on work that had been previously discussed by other students in the conference. The work was the primary basis for prompting further response.



*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, students could be asked to explain, demonstrate, and discuss the mathematical principles relevant to their sculptures. These included the design of their blueprints, executing the production of

their three-dimensional work, or elaborating on the process or meaning of their or their peers' final work. Students then would also be asked to relate what they had learned to other paths of inquiry. For instance, students would be asked not only to demonstrate how scale was applied in their sculpture, but to explain what the scale would be if their sculpture of a building were the size of the school, or how scale related to the art making in general.

In the final minutes of the protocol, students were excused from the room and then teacher reflected on what had happened during the portfolio conference. The facilitator asked the teacher how the portfolio conference demonstrated arts integration learning, if there was anything that surprised them, and if they believed an outside viewer could determine who was the pre-designated High-Average-Low performing student.

The teachers and students' responses were transcribed and then scored by offsite researchers using the below scale. Examples of transcripts and scores follow.

- NA – not applicable (if a question is not asked)
- NR - not relevant to the question
- 1.0-1.5 – Single Dimensional: Generic Response –diffuse, highly, unelaborated general statements (like “I don’t know” or “It was fun”) that lacked detail
- 2.0-2.5 – Multiple Single Dimensions: Several Concrete Concepts; Some Detail, but lacking relational thinking, cohesion, and cause and effect statements
- 3.0-3.5 – Coordination of Dimensions: Detailed Concepts and Explanation and a focus on relationships and causal links among the elements of the response
- 4.0-4.5 – Systemic Reflective Understanding: Evidence of Comprehensive Understanding that includes a demonstrated systematic knowledge and understanding of concepts, processes and other examples described and their persuasive description of the links and associations among all factors described.

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

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 and 15% mean, and 35% happy. I am whole."

Facilitator: *[Laughs.]* I like the proportion there. Now, why was 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 personal identity in terms or proportion of personality traits toward the proportion of motions in their "dance routines" that add personality to choreography they created. The relatively high rating of this strand of dialogue is based on the ability to articulate how a mathematical concept can be used in several different ways.

The second student elaborates further:

Facilitator: ... All right, Student C, what have you got?

Student C: We had to do -- first of all we had a group, and we picked, like, a theme for our group. My group's theme was respect and love. And we had to do a dance that had fractions in it, like for our locomotor movement was one-fourth of the [dance]. ...And then it would be eight counts, because we had to put in counts also. The whole dance had to be 32 counts.

Facilitator: Oh. So if you do eight counts of locomotor motion, what is the fraction of time that you're going to be spending on your locomotor? What part of the time?

Student C: Maybe like four minutes.

Facilitator: Four minutes out of how much time?

Student C: I'm not sure... Like 30 or maybe like 45 [minutes].

Rater Remarks: *2.0 Student C seems to understand fractions to a certain extent yet she confuses  $1/8$  with 4 minutes ( $1/8$  of 32 minutes), and 8 locomotor parts within 32 parts ( $1/4$ ). She cannot seem to express time in fractions at this point without mixing up concepts of minutes, hours, etc with her calculations. Were there no confusion among these concepts, and her concept of fractions remained stable, 4 minutes could have been clearly connected with its determination of being  $1/8^{\text{th}}$  of the 32 minutes.*

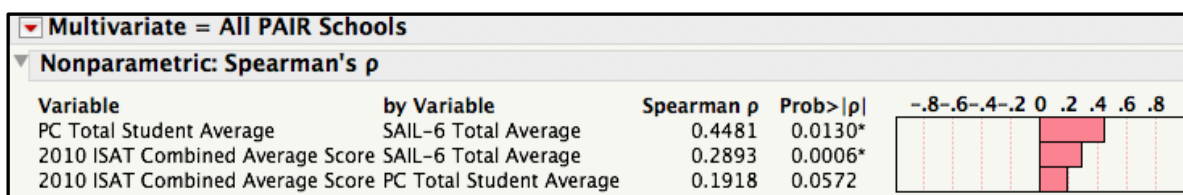
In this instance, the arts integration portfolio protocol provides opportunity for Student C to demonstrate her understanding of the computational aspects of mathematics. The

ability to articulate math concepts and make calculations that demonstrate the concepts proved difficulty for this student.

Ratings of student performance assessment tasks, as indicated by this excerpt from the Portfolio Conference transcript and scoring procedures administered only to the PAIR treatment schools, became the basis for analyzing student reflective understanding of arts integration by the end of the PAIR project.

The table below shows that, when measuring the degree of correlation between the SAIL interview, PAIR Portfolio Conference and student state test scores, the SAIL and Portfolio Conference together are highly significant predictors of 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.

**Degree of Correlation among Student Portfolio Conference Ratings,  
Student SAIL Interview Response Ratings, and ISAT Combined Academic Test  
Scores**  
(Final Report 3E-5 Figure 3)



*Putting a puzzle together: Mapping the causal links among teacher professional development, student academic, and student arts integration learning outcome factors*

The SAIL and Portfolio Conference became the two principal alternative measures of students' arts integration learning tested for their association with standardized test scores. They measured types of knowledge not covered by standardized tests, in addition to measuring student understanding of how processes and concepts essential to specific content areas (art and academic) can relate to one another when solving problems.

Researchers analyzed the SAIL and Portfolio Conference data alongside other measures to determine how teacher performance data might be linked causally to arts integration outcomes. Having gathered professional development, teacher development outcomes, student art work, student performance outcomes, each strand of the 'Rubrics Cube' could be assessed to determine the strength of their inter-relationships along the 'chain of evidence' for determine the impact of PAIR program.



In total, the variables representing multiple independent forms of teacher and students learning in the PAIR program are as follows:

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 PAIL Work Sample Ratings
- A-VI: Combine 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
- B-III: Student Portfolio Conference Interview Response Ratings
- B-IV: Student Combined ISAT Academic Performance Ratings

Although the SAIL and Portfolio Conference data are the primary arts integration student learning measures in PAIR, additional data, particularly teacher performance data, were key to this study. Several of the above variables are briefly described below.

A-I: Teacher Key Effect Ratings: Co-Investigator Dr. Gail Burnaford used qualitative and quantitative data to create four categories of teacher impact “based on whether these teachers demonstrated significant differences from their peers on specific variables, as reported on the surveys, the coded comments from portfolio conferences, the coded open-ended responses on the surveys, and the reported pedagogy identified on student work labels” (Burnaford 2010). The following effects were tested to show possible relationship to student achievement.

- 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 collected student work and reflected on their own teaching practice emphasized to make that student work.
- The Fourth Grade Effect: Three Years of Professional Development and Implementation Experience Deepens Teacher Understanding  
As the program expanded grade coverage across a three-year period, teachers in the 4<sup>th</sup> grade had three years of professional development opportunities and implementation experience in comparison with 6th grade teachers, who had only one year (although professional development sessions were offered to all teachers, 6th grade teachers rarely attended).

- The School-Wide Effect: School-wide Participation Enhances Program Implementation  
This school had strong and consistent attendance at professional development sessions by not only teachers partnering with teaching artists, but also lead arts teachers at the school.

#### Survey-based Outcome Variables:

A-IV and B-I: The Teacher Year End Curriculum Survey (YECS) survey and Student Surveys were both designed specifically for this program. The questions were based on themes presented in professional development including collaboration (either with teaching artist or students with each other), participation in art making, integrating arts with academics, reflection and dissemination. Similar to the SAIL and Portfolio Conferences, these surveys were tested for at least a year. Their questions helped formulate questions in the Portfolio Conferences to reflect program goals more consistently across tools.

#### A-V: Classroom PAIL Work Sample Ratings:

Teachers collected student work throughout the project and also collected academic-focused work from classes they, not the teaching artist, 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 used student work as data to rate teacher development outcomes.

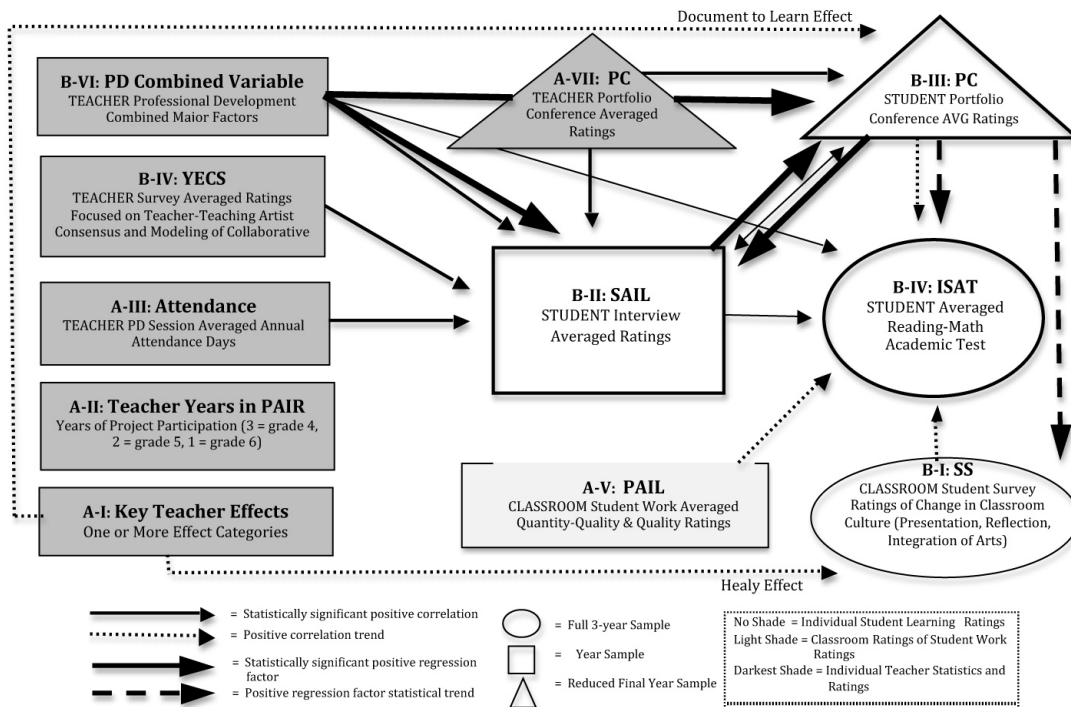
#### A-VI: Combine Teacher Professional Development Ratings:

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.

In order for the data to be a valid representation of the program — thus giving researchers an ability to make judgment of causal impact, the outcome variables were assessed in relation to one another. A series of correlation and regression tests performed are represented in the figure below to provide a perspective of the relative strengths of the program factors (on the left-center) on student arts learning (center-right) or academic learning (right). As indicated the dotted arrows indicate significant positive statistical trends in the data, the solid lines indicate statistical significance, and the thickest lines indicate the most powerfully predictive factors in relation among all other variables including student demographics such as gender, ethnicity, and family income.

## A Correlation-Regression Map of PAIR Combined Teacher-Student Outcome Intersection Factors

(Final report 3G-5 SUMMARY Figure 5)



In the analysis of the variables depicted above, researchers concluded the following:

- The pattern and degree of correlation and regression factor analysis 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 establishes causal evidence for the PAIR program's impact on student learning.
- The pattern and degree of correlation and regression factor analysis demonstrate that both individual student arts integration assessment outcomes most directly predict academic test results controlling for achievement, gender, family income, prior academic achievement ratings, or ethnicity.

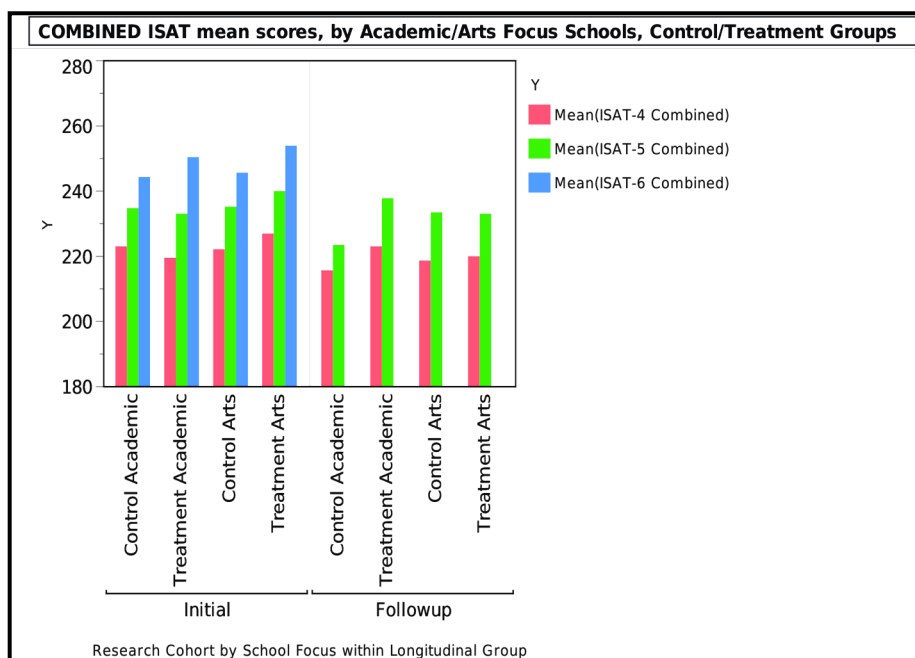
When analyzing all these factors statistically, researchers questioned how data can be looked at, quantified if possible, and if such variables could be averaged together. Researchers paid close attention to how qualitative responses were scored to ensure consistent and valid results. In the end, we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

## *Creating an Optimal Condition in Schools: Equity and Excellence in Student Learning Outcomes*

As demonstrated from the SAIL and Portfolio Conference assessments, students from treatment schools with a focus in the arts, scored significantly higher or on par on these alternative assessments than any other type of school. Students from treatment schools with a focus on academics frequently scored the next highest, with the students from the control schools scoring lower (on the SAIL). Additionally in the treatment schools, pre-classifications of students according to 3<sup>rd</sup> grade High Average Low test scores did not determine how they performed on the SAIL by the time they were in 6<sup>th</sup> grade. *These classifications did hold, however, in the control schools.*

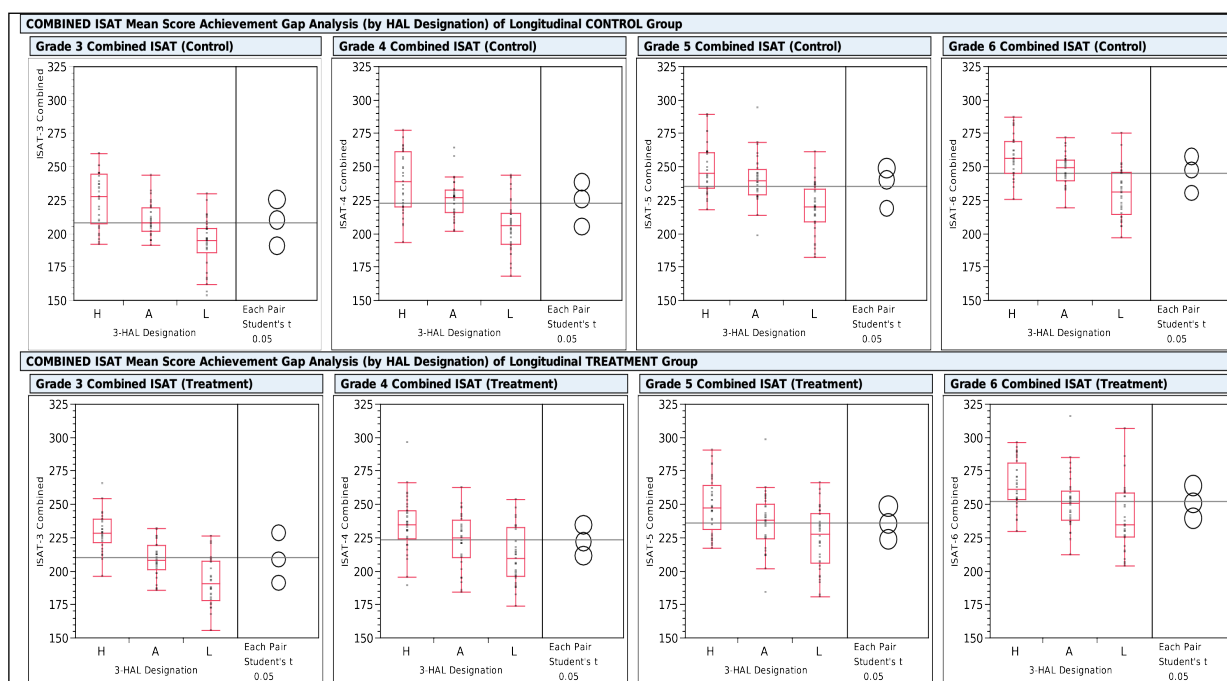
The suggestions raised from these alternative assessments remain true when looking at student test score data. The longitudinal cohort in the treatment schools not only outperformed the control schools on their averaged math-reading scores, students from *the treatment schools with an arts focus* particularly exceeded. The results still hold for the students following the primary longitudinal cohort — those students who participated in the program for two years instead of three. The difference is sustained, if not widened between the treatment school student scores and the control students. Researchers found that for those with only one year of experience in the program, there are no significant gains made over the control students' scores.

### **Control-Treatment PAIR Focus School (Arts vs. Academic) Comparisons of ISAT Mean Scores, PAIR Initial Longitudinal Cohorts, Grades 4-6 (Final report 2D Figure 1)**



Furthermore, in the control schools, low-scoring students never approach middle or high scoring students even when they reach the 6<sup>th</sup> grade. Yet in the treatment schools, the initially low-scoring students gradually approach those higher scores. In fact in the 6<sup>th</sup> grade, several initially low-scoring students had very highest scores of any control or treatment student. This phenomenon of “closing the achievement gap” also did not sacrifice student scores that were initially high.

### Control-Treatment PAIR Schools Achievement Gap Box Plot Analysis of HAL ISAT Combined Academic Scores, Grades 3-6 (Final report 2E Figure 2)



Based on these performance measures, the coherency of the data, and the variables collected, researchers can begin to make specific recommendations for future arts-based interventions and specific changes in future research tools. As the data displays demonstrate, student scores do not exhibit improvements if the intervention only lasts one year. So this study makes a strong case for programs that give students at least two years of exposure to this arts-based intervention. Also, because student scores are correlated with long term, high quality teacher participation, we can also make the case that teachers necessarily participate in professional development, document, reflect and collaborate with the same teaching artist in these types of projects for at least two years.

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. Multivariable analysis provides a means of understanding how that how that impact takes place. In both examples below, a teacher observes the evolving impact of arts integration on her students, in ways that are mirrored by the growing statistical impact of PAIR on arts and academic learning in relation to the control schools.

Well, the original goal was to see how the arts itself affected the students' abilities as far as their writing, and how the different arts, for example, music, because we had a percussionist and a playwright, would enhance the child's development of characters and stories. The first year was pretty new. We still had to develop some skills as far as what the artists were teaching the children and my own understanding of what it was. This year I think it's been...I've seen...to get the kids to actually think more critically. Last year it was okay, okay, it's going through the machinations, for the most part. This is what I'm supposed to do, this is what we did (4<sup>th</sup> grade teacher).

Well, I think these students are much better at relating it to math than last year. I think they were much more...they expressed themselves much better mathematically, which is because when I was in the room, we really stressed this is not only art. I mean, we talked about it before, but we really stressed it this time because I didn't want it to happen again where they didn't see the value of the math in the project (5<sup>th</sup> grade teacher).

### *Concluding Statements*

For many years, arts learning organizations and researchers have been unprepared to discuss the statistically determined causal links among the various discrete elements of arts-based interventions in AEMDD programs. Often shying away from facing the 'burden of proof' for the "success" of arts integration programs, the stakeholder of arts integration programs in schools are left in the lurch between describing the effectiveness of program implementation and changing test scores without being able to specifically link teacher and student outcomes statistically. This project demonstrates the results of a methodology that demands discussion and accountability for outcomes on a variety of program, teacher and student levels. The 'Rubrics Cube' multivariate analysis methods used in PAIR have been invaluable for studying complex learning environments in the past. The results discussed here show that these methods can determine causal links among teacher professional development and student learning outcomes to deepen a school community's understanding about the spectrum of arts integration effect and how these interventions impact the quality of school culture and achievement.

Additional arts learning assessments are a necessary practice for determining the success of arts-based interventions in education. They provide the basis for determining substantive causal links that progress from program factors to teacher and student learning outcomes. In the case of the PAIR program, the SAIL and Portfolio Conference provided two different, validated measures of arts integration learning, the second of which also served measured teacher performance and served as teacher professional development.

In order for assessments to be valid and reliable, researchers and practitioners (including program staff, artists, teachers) must dialogue on a continual basis, to make sure that the

assessments meet the goals of the program, and reflect the goals in a consistent way amongst different tools. This simultaneously builds an organizational capacity to understand research to improve programming. Without this collaborative research practice, data frequently yields limited findings or is unusable.

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 — there is now statistical evidence to claim that arts integration programs improve academic performance as well as close the achievement gap (from slide). Students in the treatment schools not only outperformed students in control schools on the SAILS and standardized tests after three years of the program, but previously low scoring students were approaching higher scoring students, whereas students in the control schools remains stratified. Additionally, viewers of the Portfolio conferences could not reliably distinguish between the High, Average and Low students on the basis of the ability to demonstrate, critique or reflect on their arts integration work. This further demonstrates that when students are given to the opportunity to demonstrate their arts integration learning in more ways, schools can claim evidence of equity and excellence provide substantial, high quality arts integrated instruction. That arts integration assessments also are highly predictive of state standardized tests of reading and math suggests that arts integration teaching and learning provides a school culture of high achievement that also closes the gap between the formerly low achieving students. That the arts integration treatment schools outperformed conventional arts or academic focused schools suggest that arts integration is an optimal condition for the impact of arts education on academic achievement.

The methodology and tools presented are not intended as definitive; future studies will likely adapt them to suit the project under study. But the groundwork they lay for art education, and an organization using them, can be widely built upon. Future research studies at CAPE will build on this multivariate methodology to continue the study of equity and excellence in student learning, particularly in schools that will support a wider spectrums of portfolio assessment and arts integrations practices in International Baccalaureate schools, Arts magnet schools, or any other neighborhood school that will take serious the impact of documenting and assessing arts integration learning for all students.