

**Chicago Arts Partnerships in Education's
Chicago Public Schools
International Baccalaureate – Teaching Artist Project
(IB-TAP)**

**Principal Investigator's
Student Learning Outcomes
Statistical Analysis Report**

**Submitted by
Dr. Lawrence Scripp, Co-Principal Investigator
Center for Music and the Arts in Education (CMAIE)**

**Fred Sienkiewicz, Devin Ulibarri, Sarah Sutherland, Josh Gilbert,
Research Assistants**

October 27, 2014

Abstract/Executive Summary

The Chicago Arts Partnerships in Education (CAPE) International Baccalaureate–Teacher Artist Project (IB-TAP), supported by a four-year Arts in Education Model Development and Dissemination (AEMDD grant, initially engaged ten International Baccalaureate (IB) Middle Schools in the Chicago Public School District (CPS). Six of these IB schools incorporated 20 units supported by CAPE teaching artists in 6th, 7th and 8th grade treatment school classrooms while four Control IB Schools maintained their standard IB practices.

The purpose of the project was to see to what extent CAPE arts integration teaching and learning methods—engaged by the collaboration of CPS IB classroom teachers with CAPE teaching artists—could optimize teaching and learning in IB schools already committed to arts education, portfolio assessment, and integrative teaching and learning. As reported by Dr. Louanne Smolin, CAPE staff and teaching artists provided extensive professional development for classroom teachers through the three years of project implementation. Teacher professional development outcomes were measured through evaluation of exit interviews, classroom observations, and teacher portfolios. Results showed substantial improvement of teacher qualitative measures of teacher collaboration, planning, co-teaching, peer coaching, and student project documentation and reflection.

Student learning outcomes in the project were derived from:

1. Student Illinois State Achievement Tests (ISAT) in Math and Language Arts,
2. Student Arts/Arts Integration Performance Assessment Interviews (PAI) used for determining levels of sophistication in response to questions and tasks focused on (a) knowledge of various art forms, (b) skillful arts making processes, (c) judgment about the characteristics of high quality arts, (d) the means and impact of artistic expression, (e) Imagination in the creative artistic process, and (f) aesthetic experience in response to art works.
3. Student Arts/Arts Integration Portfolio Conference (PC) Performance Assessment Interviews used for determining levels of student sophistication in presenting their work and answering questions about their classroom work with teaching artists, the work of their

peers, and to describe how and why their IB-TAP units contribute to their classroom learning and performance on standardized academic tests.

Statistical analysis of IB-TAP classroom teacher performance outcomes focused on four measures of professional learning indicators:

1. Teacher Years in the IB-TAP project.
2. Teacher Days of Participation in IB-TAP Professional Development Sessions.
3. Teacher Arts/Arts Integration Portfolio Conference (PC) Interviews: A protocol used for determining teacher levels of understanding the principles and practices of arts integration and its impact on student performance.
4. High Quality Teacher Ratings developed by co-principal investigator Louanne Smolin and Teacher Portfolio Conference Ratings.

After controlling for outlier schools and radical changes in ISAT standardized tests in the final year of the project, longitudinal comparisons between control and treatment school student cohorts showed that that IB-TAP students improved significantly more on combined math-language state standardized tests than did the control school cohort students. Within the treatment schools cohort, stepwise regression analyses revealed that the Student Arts/Arts Integration Portfolio Conference Ratings were the primary predictor of accelerated ISAT academic performance, outpacing all student demographics (including ethnicity, prior academic performance, and social economic status) and every teacher quality rating variable (years in study, attendance in professional development workshops, teacher portfolio conference, and high quality ratings).

Results from this study suggest that when on-going, high-quality professional development is provided to support the collaboration between arts-integration focused teaching artists and IB middle school teachers, evidence of accelerated positive academic achievement—and arts learning—results. Furthermore, these results indicate that the use of student portfolio assessment methods is the best measure for the effectiveness of arts integration teaching and learning practices. These results, however, may have been far more conclusive, had there not been (a) extensive attrition in the sample school longitudinal cohorts, (b) the need to reject outlier schools school cohorts within both the treatment and control groups, or (c) changes initiated by the state board of education that imposed radical changes in the ISAT test design, implementation, and test preparation that led to erratic test results during the final year of the project.

* * *

Section 1: Introduction to Investigative Evaluation Methods

The Chicago Arts Partnerships in Education (CAPE) International Baccalaureate–Teacher Artist Project (IB-TAP), supported by a four-year Arts in Education Model Development and Dissemination (AEMDD) grant, proposed a quasi-experimental study to determine the impact of CAPE arts integration methods on student academic and arts learning performance in International Baccalaureate (IB) Middle School in the Chicago Public School (CPS) district schools. Six IB schools forming the treatment cohort agreed to incorporate twenty CAPE units into the middle school curriculum that were created by teaching artists in collaboration with 6th, 7th and 8th grade treatment school classrooms. Four IB Schools who agreed to serve as a control group cohort maintained their standard IB practices in their classrooms throughout the three years of IB-TAP program implementation. Control and Student cohorts were determined by random selection within pools of pre-designated high (H), average (A), and low (L) academic performing students in each classroom that initially resulted in demographically and academically balanced sample groups.

As the project proceeded, Control-Treatment (C-T) School comparisons of student performance academic outcomes drew on (a) Illinois State Achievement Test results and (b) ratings of student understanding of arts and arts integration learning ratings derived from individual student Performance Assessment Interview (PAI) protocol¹ described in detail in Section 4 of this report.

Within Treatment schools, analysis to determine links between arts integration teacher professional development and student learning outcomes drew on data collected from the following teacher variables: (a) Teacher Years in the IB-TAP project, (b) Teacher Days of Participation in IB-TAP Professional Development Sessions, (c), Arts/Arts Integration Portfolio Conference (PC) Interviews² and (d) High-Quality Teacher Ratings³.

¹ PAI protocols, adapted by the Center for Music and the Arts in Education (CMAIE) for the IB-TAP project, The protocol and rating rubric are described in Section 5 of this report.

² A protocol developed by the Center for Music and the Arts in Education (CMAIE) for the IB-TAP project used for determining levels of teacher’s articulation of the principles and practices of arts integration and its impact on student performance as described in Section 5 of this report.

³ A ratings system developed by Dr. Louanne Smolin for the IB-TAP project used for determining levels of teacher effectiveness as described by Smolin in her IB-TAP co-principal investigator’s report.

The principle research questions are (a) are pre-post changes in ISAT scores in the treatment school longitudinal student cohort higher than those of the control school? and (b) What teacher or student factors best predict this result? In order to investigate these questions, the evaluation design had to be adjusted (a) to accommodate imbalances in student cohorts caused by student attrition by eliminating outlier schools from both the control and treatment school cohorts to create a fair comparison, and (b) to assess the impact of major changes in the CPS testing program that were aimed at holding students to the higher standards established by the adoption of the Common Core State Standards in the future, but temporarily created unintended baseline differences in performance between the control and treatment schools that confounded the evaluation in year 3 of the IB-TAP project.

The results that address the first evaluation question can be used as the performance measure required under the Government Performance and Results Act (GPRA) for the AEMDD grant program:

GPRA Performance Measure, Indicator 8.1.1: Percentage of participants who benefit from standards-based arts education and meet state learning standards will increase.

Performance Data: The numbers and percentages of students participating in arts models programs that demonstrate higher ELA and math achievement than those in comparison schools.

Following this introduction, the report describes the method of adapting the revised evaluation methods. Succeeding sections will present the results followed by a summary and conclusions.

* * *

Section 2: Preliminary Analysis to Determine the Validity of IB-TAP Control and Treatment School Student Standardized Test Score Comparisons

Section 2A: *The Discovery of Outlier Schools in the Original Longitudinal Comparison Sample*

Originally, the student cohorts based on randomized sampling in four control and six treatment CPS International Baccalaureate middle schools were well matched in terms of academic performance, ethnicity, gender, and social-economic status (percent of free and reduced lunch students). Due to student attrition and teacher dropouts, however, the overall longitudinal cohorts comparisons were

affected unequally, resulting in statistically significant differences in levels of Combined Math-Language Illinois State Academic Test (ISAT) ratings from the start of the project (see Figure 1 below) and therefore skewing the starting baseline Illinois State Achievement Test (ISAT) data to the extent that the comparisons were inherently unfair throughout all subsequent phases of the analysis.

This unfair longitudinal baseline year comparisons obscure the impact of IBTAP Program on International Baccalaureate CPS schools because (a) the baseline comparisons suggest a significantly different level of academic performance (246.3 compared to 231.5 ISAT scores) and (b) although the IBTAP students appear to be closing the gap by Year 2 (Year 3 comparisons will be discussed in section 2B), it is highly likely that the overall treatment school performance is significantly lower because of the original imbalance in the student cohorts.

Figure 1: Averaged Baseline (Grade 5) and Three Implementation Years (Grades 6-8) Combined Math and Language Arts ISAT scores for All Control-Treatment Chicago Public School IB Schools.

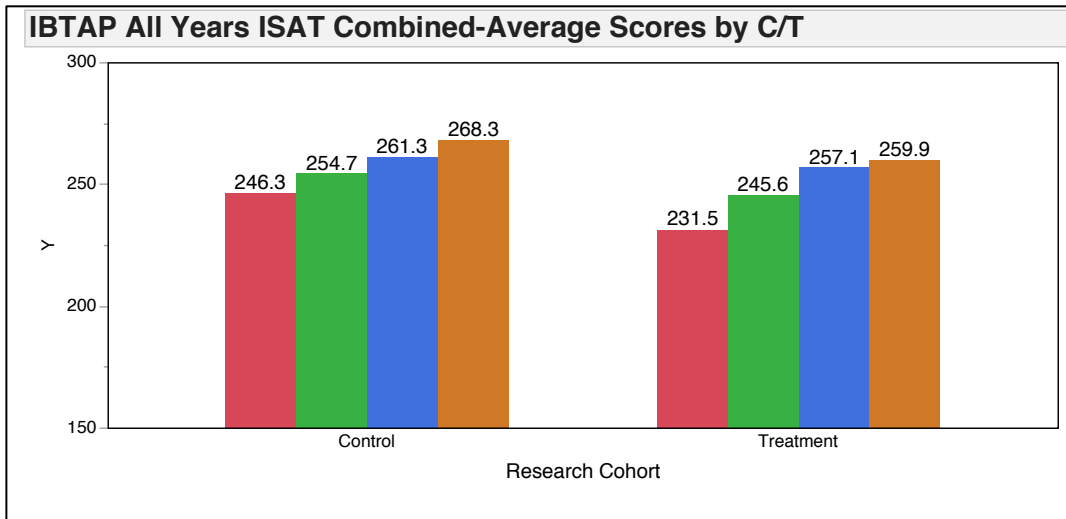
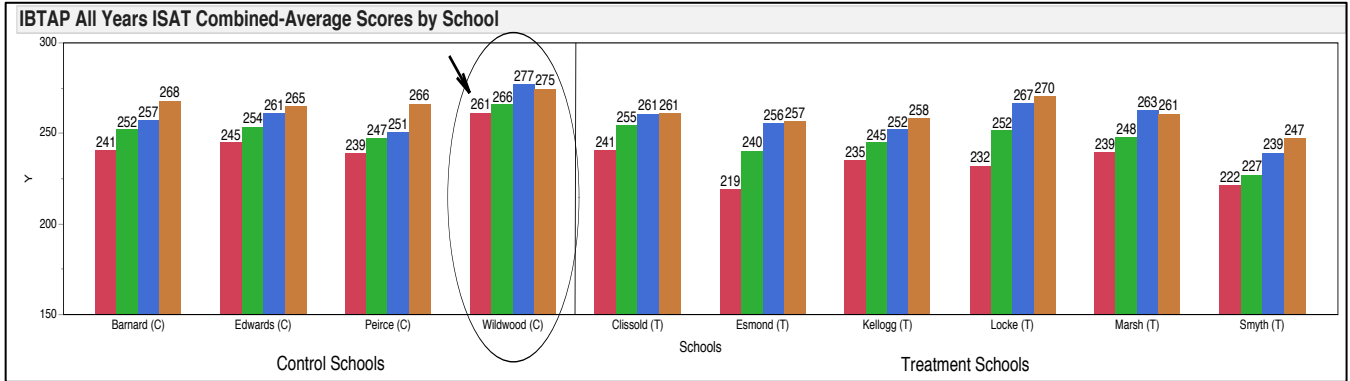


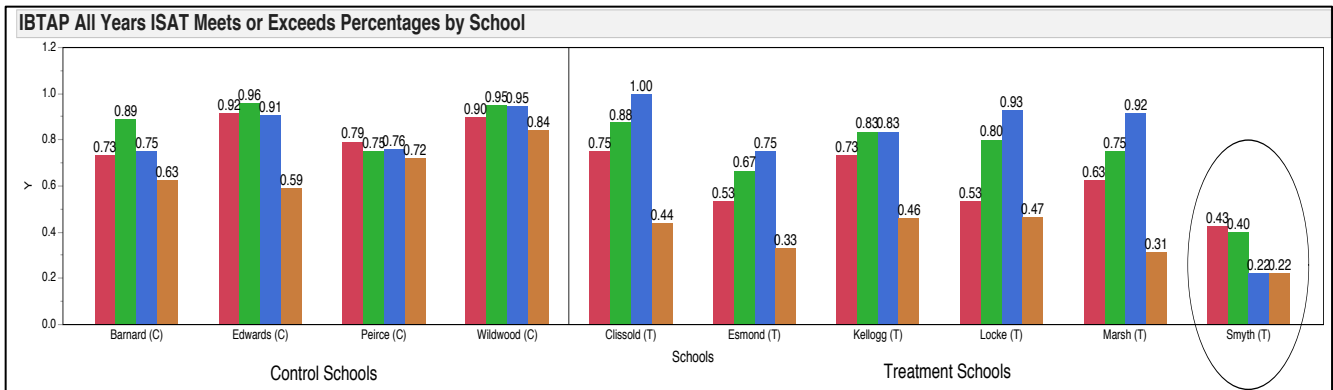
Figure 2 below reveals that that Wildwood control school (fourth school from the left) is an outlier compared with any of the other control or treatment schools particularly with regard to its baseline academic performance (261 compared to the next highest score of 245). Thus the one individual school outlier represents the major cause of the baseline data distortion.

Figure 2: Comparison of All IB Individual School ISAT performance



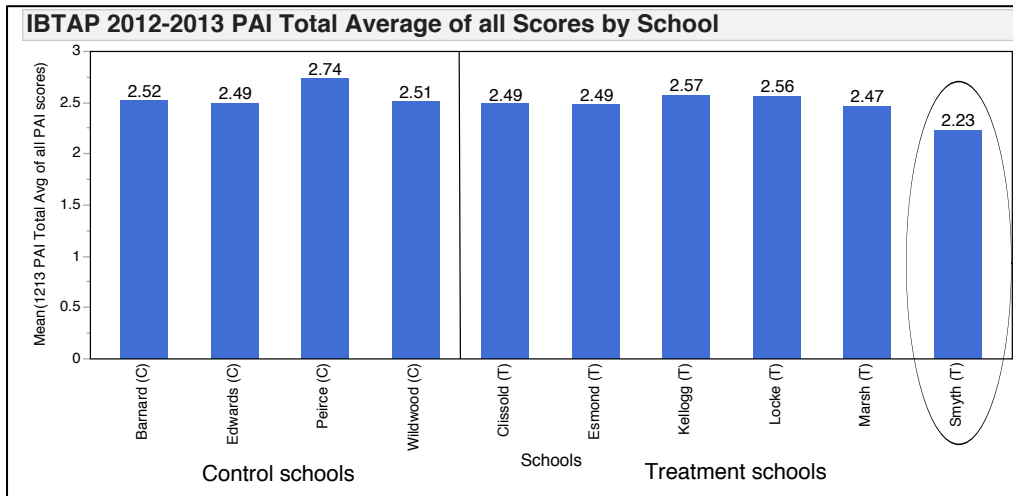
The Smyth school (last school on the right) also revealed itself to be an outlier treatment school due to several factors. First, Smyth’s ISAT average scores turned out to be skewed by a small percentage of high scoring students. Second, as revealed in Figure 3, CPS district benchmark ratings show that the Smyth school student cohort (last school on the right) had by far the lowest percentage of students who “meet or exceed” (MEX) CPS academic standards compared to any of the other control or treatment schools (.10 below the next lowest school baseline statistics) in all four years of the project. As indicated in the chart, this gap in district benchmark standards only increased over time (Year 1: -.17; Year 2: -.53) resulting in the Smyth IB being regarded as failing school in each of the four years of the project.

Figure 3: Comparison of Percentage of Students who “Meet or Exceed” (MEX) CPS district academic performance standards



Smyth also disqualified itself as a treatment school because of its lack of faculty participation in the project professional development program and the extremely subpar level of implementation of IB-TAP practices that resulted in this school being, by far, the lowest-rated school in terms of student arts and arts integration performance assessment scores. As indicated in Figure 4, the IB-TAP Performance Assessment Arts Integration (PAI) scores are much lower in Smyth than any other school: the Smyth student averaged PAI score (2.23) is 0.31 points below the average score of the 9 other treatment and control schools (2.54), which is over 2 standard deviations below the mean.

Figure 4: Comparison of All School Student Cohorts IBTAP Arts/Arts Integration Performance Assessment (PAI) Ratings



Section 2A Conclusion: *Removing Two Outliers Provides a Better Basis for Control-Treatment School Comparisons*

The Wildwood control school data, principally because of its disproportionately high baseline academic scores, skewed the control school cohort averages to the point where a fair comparison cannot be made between the control and treatment schools during the course of the project.

The Smyth treatment school data, principally because of its disproportionately low percentage of CPS district students and its failure to adopt IBTAP treatment school practices, invalidated Control-Treatment school comparisons. In effect, this school functioned more like a school like a control schools, whereas the Wildwood school, because of its own initiatives with arts integration, functioned more like a treatment school.

Because of the discovery of imbalances and anomalies in the longitudinal cohort samples that developed over time, the Wildwood control and the Smyth treatment schools will now be eliminated from all the data analyses.

Figures 5A and 5B demonstrate how the removal of the two outlier schools has significantly improved the balances in the baseline data. That is, the averaged baseline ISAT score gap between the first columns on the left in the control and treatment school data display has been reduced by nearly 45.3% (from 14.8 points in Figure 5A to 8.1 points in Figure 5B).

Figure 5A: ISAT Data Display BEFORE Removing the Outlier Schools (C-T School Baseline Year 1 Difference Score = 14.8 points):

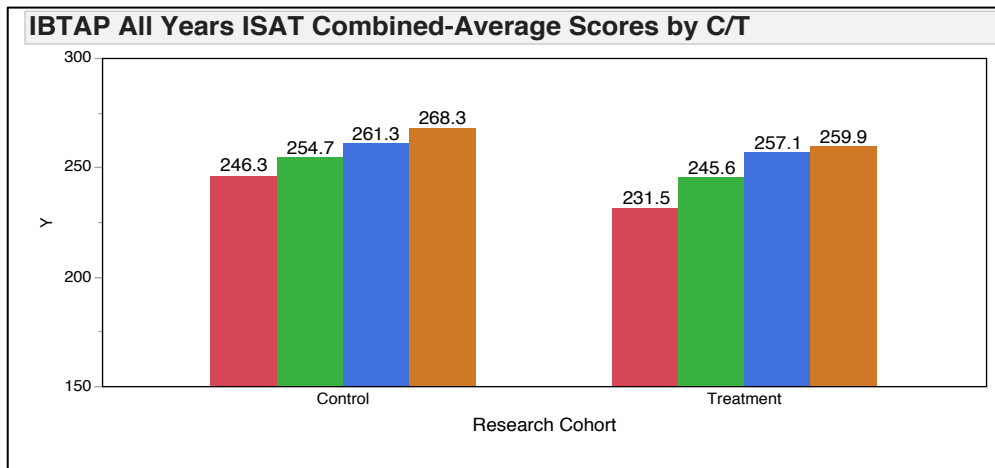
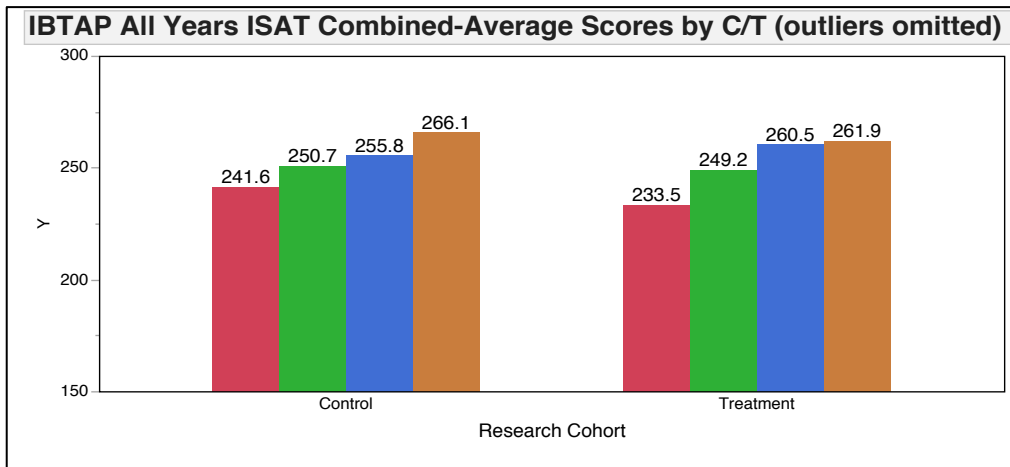
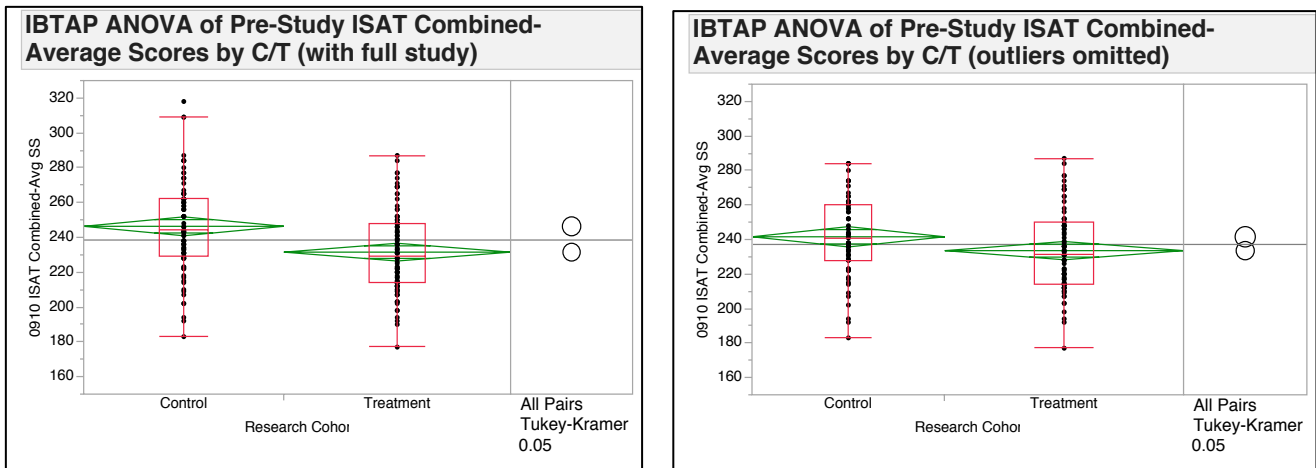


Figure 5B: ISAT Data Display AFTER Removing the Outlier Schools (C-T School Baseline Year 1 Difference Score = 8.1 points):



Removing the outlier schools from both control and treatment school cohorts provides a much fairer baseline comparison. As seen in Figures 5C and 5D the difference between the ISAT baseline gap BEFORE outliers were removed (mean difference of 14.83, t ratio = 3.93, p = .0001) and the baseline gap AFTER outliers were removed (mean difference of 8.03, t ratio = 2.01, p = .05) is relatively much less pronounced.

Figure 5C-D: Two ANOVA Data Displays Comparing Statistical Differences between Year 1 Baseline Scores BEFORE (full study) and AFTER (outliers removed) Eliminating the Outlier Schools.



It was the view of the research teams that, by removing these two outlier schools, data analysis could proceed providing there were no other invalidating factors present in the study. Unfortunately, the issues of ISAT testing changes in the final year of the project limited the scope of the ISAT data analysis significantly as well.

* * *

Section 2B: *Changes in Standardized Tests Puts Final Year Results in Doubt*

Signs of anomalous ISAT results in Year 3

As can be observed from the data displayed in Figure 5B above that, after two years of dramatic improvement, that the ISAT scores for the Treatment Schools did not improve at all in the final year of the project. To find out why the improvement ceased in the Treatment schools, the research team found that that the Illinois State legislature had approved changes in ISAT tests that went into effect during the third year of the project. These new tests were intended to be more closely aligned with Common Core standards in language arts and math, yet the CPS did not ensure uniform training for

teachers to prepare students for the new test nor were the criteria for CPS benchmark standards known to the students. Under these conditions it became impossible to determine what effect these new policies would have on CPS schools.

For IB-TAP school comparisons between Years 2 and 3 of project implementation showed that the Control Schools benefitted disproportionately from the new one-time “transitional test.” The schools that were not able to provide adequate teacher training for the new tests, not surprisingly, tended not to improve from Year 2 to Year 3 results as seen in Figure 6.

Figure 6: IBTAP All Years ISAT Combined-Averaged Scores by School (without Outliers)

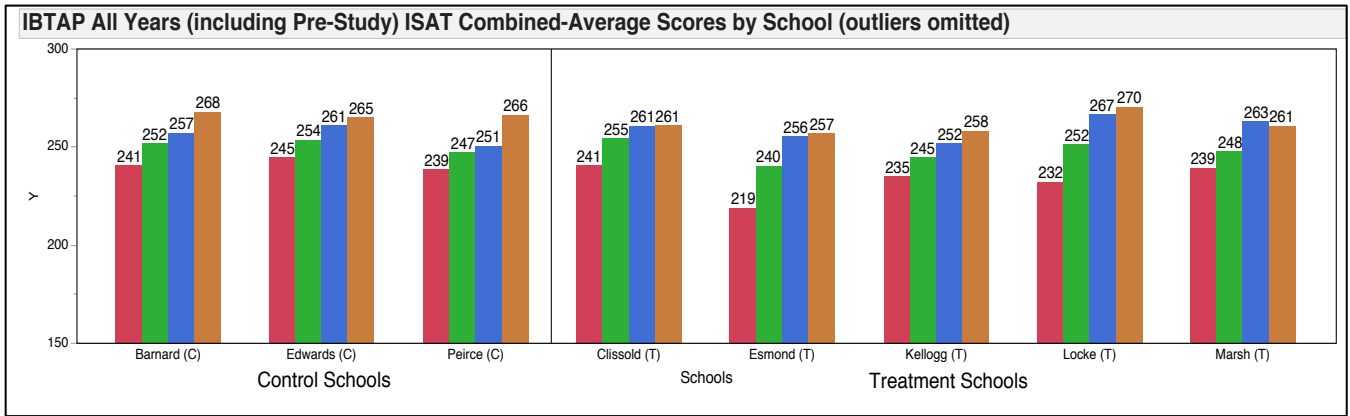


Figure 6 above disguises radical effect the new ISAT test had on year the final year of the project student academic test data.

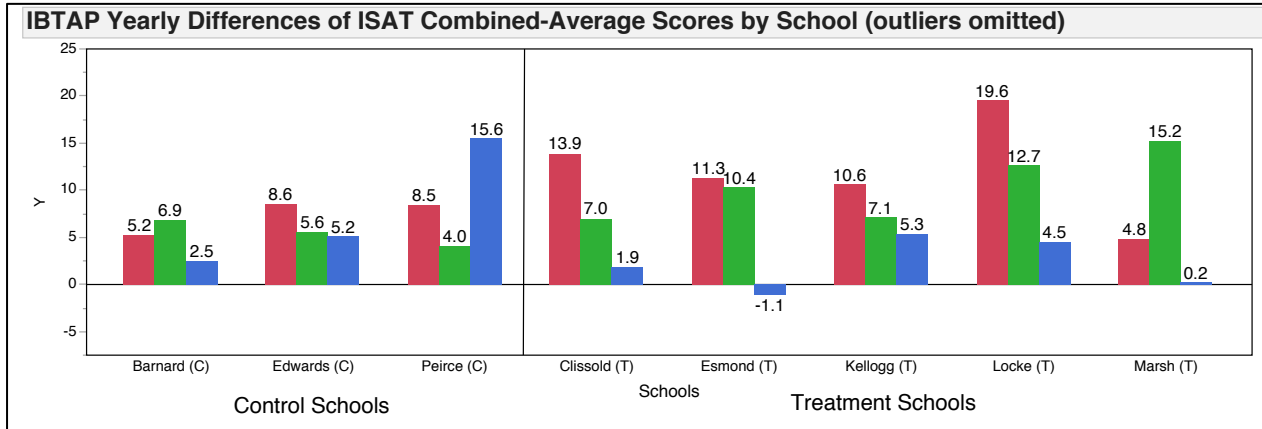
Figure 7 below reveals the dramatic effect of the new “transitional” ISAT test (which resulted in uneven teacher preparation for the test across both the control and treatment schools) on student performance. Annual differences of ISAT scores over the three phases of the IB-TAP project implementation have completely skewed the pattern of development in all schools prior to the new test.

See differences in:

- Column 1 (red): differences between baseline and implementation Year 1 *favoring the treatment schools;*
- Column 2 (green): differences between implementation Year 1 and Year 2 *again favoring the treatment schools though the degree of change is less;*

- Column 3 (blue): differences between implementation Year 2 and Year 3. *Test scores appear to violate the pattern of change of the previous two years randomly.*

Figure 7: C-T Comparison of Individual School ISAT “Degree of Improvement” Scores



Control and Treatment Schools Comparisons of Differences between Year 2 and Year 3 Project Implementation ISAT Scores demonstrate the radical difference of the new ISAT “one-time transitional’ test in the final year of the project. Differences in the final year of the project are far more random than the pattern established in the first two years of the project.

Furthermore, analysis of the impact of new benchmark formulas applied to the ISAT scores in Figure 8 show an extremely pronounced depression of the percentage of students “meeting or exceeding” CPS performance benchmarks for the Treatment Schools in Year 3 compared to Year 2 student ratings.

Figure 8: Contrasting Percentage of Students Who “Meet or Exceed” (MEX) CPS Benchmark Standards in Years 2-3 of IB-TAP

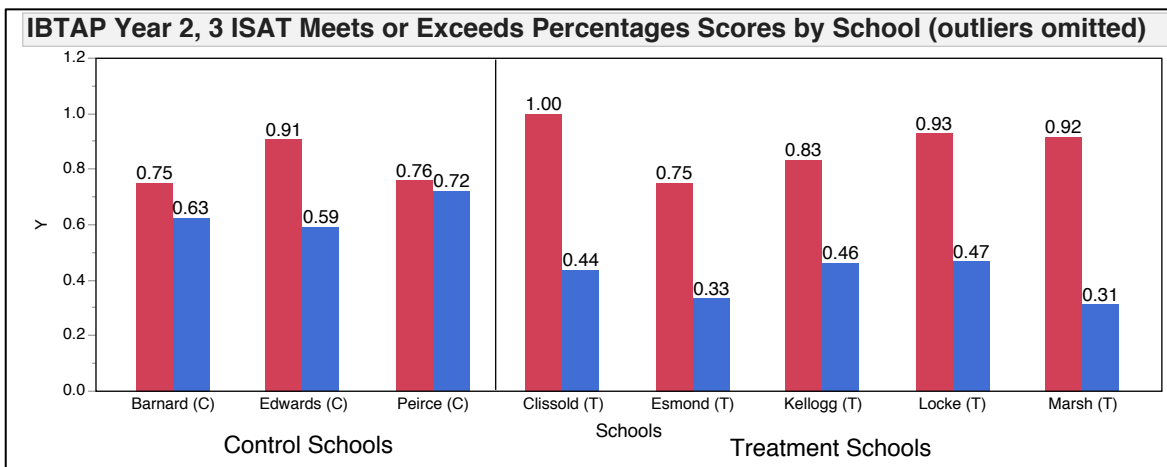
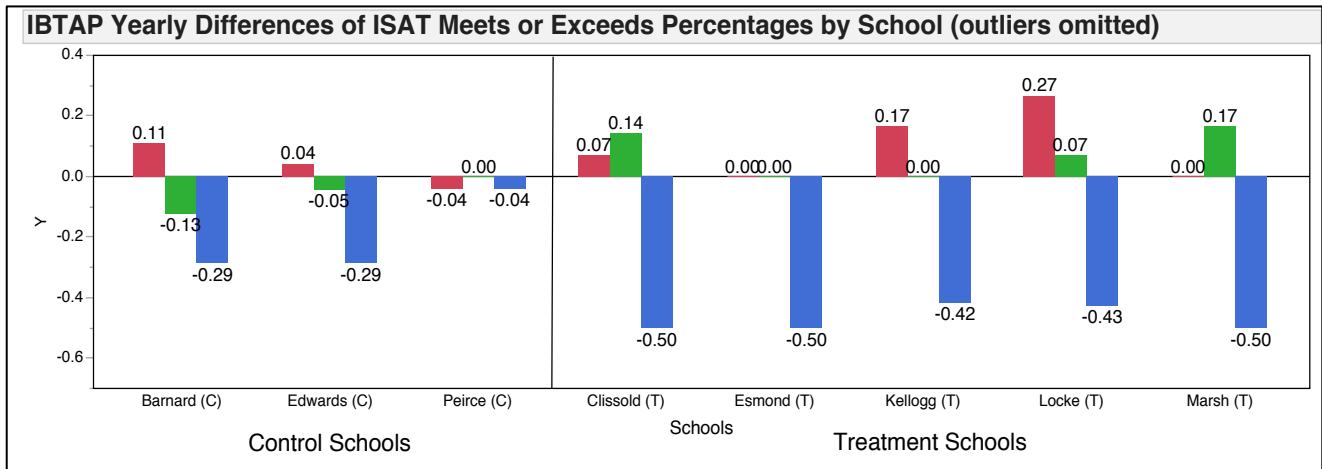


Figure 9 illustrates how differences (gain-loss) in CPS MEX benchmark scores based on the new ISAT test results muddle all previous data trends making it impossible to judge the pattern of development of any particular school as shown most dramatically in the data display below.

Figure 9: Pattern of gain-loss differences of ISAT based MEX benchmark scores



Furthermore, Table 1 demonstrates that differences between the old and new test do not stand up to close analysis. Looking at the highlighted columns, the differences of ISATs and MEXs do not reflect a coherent assessment of academic performance.

Table 1: Summary of ISAT / “Transitional’ ISAT and CPS %Meet or Exceed Benchmarks Scores

School	Year 2 ISAT avg	Year 3 ISAT avg	Year 3 – Year 2 (diff)	Year 2 MEX avg	Year 3 MEX avg	Year 3 – Year 2 (diff)
Barnard (C)	257	268	+11	0.75	0.63	-0.12
Edwards (C)	261	265	+4	0.91	0.59	-0.32
Peirce (C)	251	266	+15	0.76	0.72	-0.04
Clissold (T)	261	261	0	1.00	0.44	-0.56
Esmond (T)	256	257	+1	0.75	0.33	-0.42
Kellogg (T)	252	258	+6	0.83	0.46	-0.37
Locke (T)	267	270	+3	0.93	0.47	-0.46
Marsh (T)	263	261	-2	0.92	0.31	-0.61

Section 2B Conclusion: The third year ISAT scores, when compared to any of the patterns of growth in the first two years, are now considered anomalous and, as is the case with the two outlier schools, will not be included Control-Treatment comparisons in the following analyses.

In the case of “treatment school only” data analysis that follows, however, the outlier schools and Year 3 ISAT scores will not be omitted because the issue of fair Control-Treatment school comparisons is not relevant when arts learning assessments or ISAT test scores are only administered within the treatment schools student cohort.

* * *

Section 3: IB-TAP Adjusted Student Academic Performance Statistical Results

Section 3A: Student ISAT Outcomes Corrected by Omitting Outliers and Disregarding the Final Year “Transitional” Test Results

Looking at Pre-Study (Baseline), Year 1 and Year 2, the ISAT scores of both Control and Treatment increase each year. However, by the end of the project the “Baseline to Year 2” margin of improvement for the Treatment Schools almost twice as much as the Control Schools.

Figure 10: C-T School ISAT Combined Score Comparisons from Baseline to Year 2 of IB-TAP Project Implementation

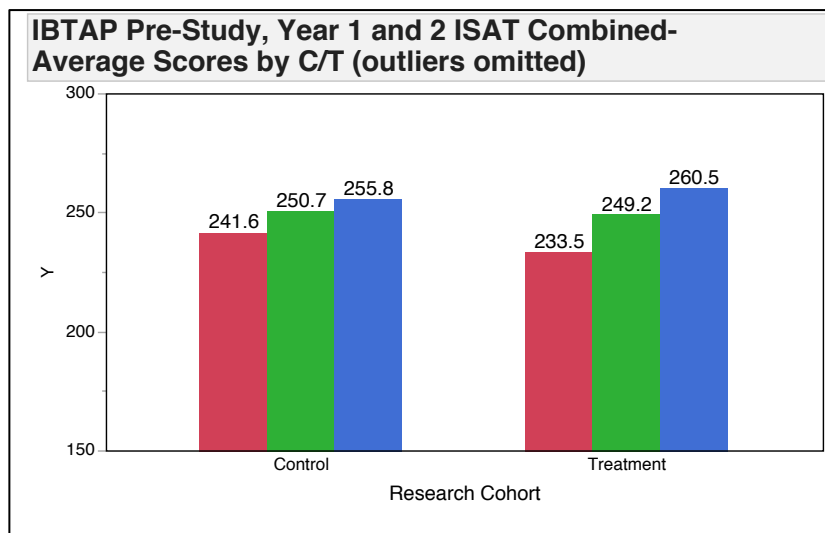


Figure 11 shows how ISAT scores were distributed according to categories of student academic achievement designated before the study started. HAL analysis shows that all three types of students (high, average, and low pre-designated students) improved more in the treatment than the control schools, and that the most highly rated (H) pre-designated academic achievers in the study benefitted most from the IB-TAP program.

Figure 11: C-T Comparison of ISAT Scores Baseline to Year 2 of IB-TAP According to Pre-Designated Levels of Academic Achievement

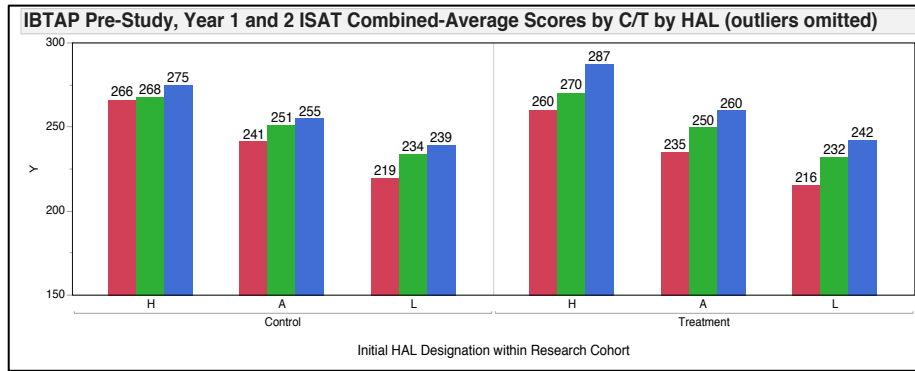
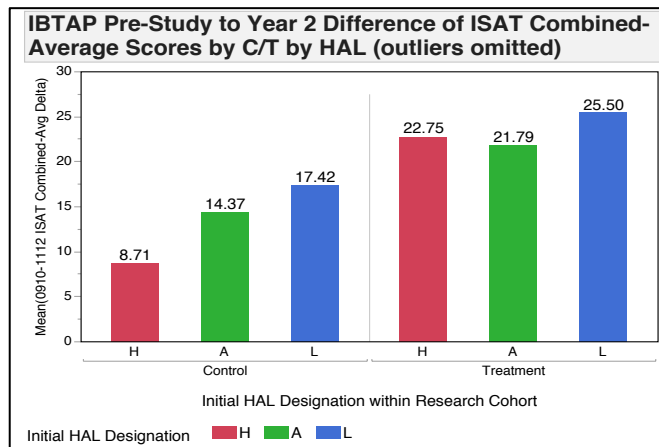


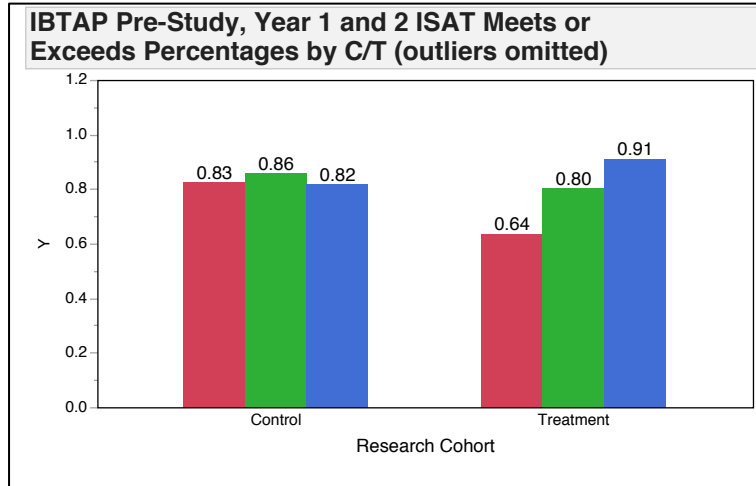
Figure 12 shows the extent to which the pre-designated HAL treatment school students improved their ISAT scores at a greater rate than did the similarly categorized control group students. When broken down by HAL, this figures shows that in the Control schools, the High (H) students had relatively low improvement with 8.7, followed by Average (A) students with 14.4, and the Low (L) students improved the most with 17.4. The treatment schools, however, show all three groups improving between 21.8 and 25.5, showing uniform improvement at a much higher level than the Control group students.

Figure 12: C-T Comparison of ISAT Gain Scores from Baseline to Year 2 of IB-TAP According to Pre-Designated HAL Levels of Academic Achievement



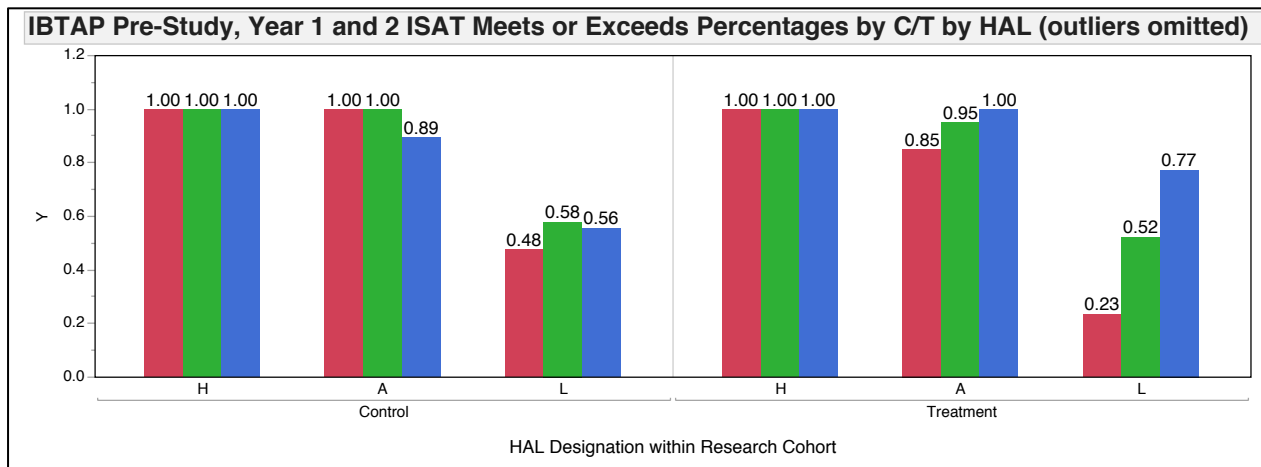
Looking at the percentage of students who met or exceeded the CPS benchmarks for ISAT score shows the treatment school MEXs improved at a far greater rate than did the control school students. The MEXs for Control schools stay hovering between 82% and 86%, while those of the Treatment schools start at 64% and increases to 91%.

Figure 13: C-T Comparison of Percentage of Students Who Meet or Exceed CPS ISAT Benchmarks from Baseline to Year 2 of Project Completion.



When looking at MEXs by HAL ratings in Figure 14 below, the Control starts at or above Treatment with respect to the average and low rated academic achievers. However, Treatment IB-TAP students perform at or above the Control students in all three HAL categories by the end of Year 2 of the project.

Figure 14: C-T Comparison of Student MEX percentages by Year 2 of IB-TAP



Section 3 Conclusions: Adjusted Control-Treatment School comparisons provide clear evidence for the success of the IB-TAP project in terms of ISAT combined academic achievement scores, rate of improvement, and in terms of serving all three categories of HAL pre-designated academic achievers when compared to the control schools.

The next sections will look for statistical relationships with ISAT scores in order to determine what student learning or teacher quality ratings accounted most for the differences in academic achievement.

* * *

Section 4: Analysis of Arts and Arts Integration Student Performance Assessment Interview (PAI) Ratings

Section 4A: *The PAI Data Collection and Analysis Methods*

Student Arts/Arts Integration Performance Assessment Interviews (PAI): A protocol developed and adapted for IB-TAP by Dr. Lawrence Scripp was used for determining levels of sophistication in answering questions about art, art making, and the experience of responding to art. The following questions, interspersed by responding to examples of art, were posed during 45-minute video recorded interviews in the final year of IB-TAP:

- Knowledge of Art Forms:
Arts questions: What is art (music, dance, drama)? Please give an example of high quality arts? What makes this example high quality? What makes this artwork high quality?
Arts Integration question: How is high quality music and visual art similar to or different from high quality literature or thinking in language arts and/or math?
- Skillful Arts Making Processes:
Arts questions: What are useful skills and processes for creating art? Why is it important to learn these skills and processes?
Arts Integration question: How are creative processes in music and the visual arts similar to or different from language and math?
- Artistic Judgment:

Arts questions: How do you know when you have finished successfully creating or performing a piece of art? How do you know when you have successfully finished creating a work of art?

Arts Integration question: How are judgments about completing music or visual art similar to or different from completing language arts writing or solving mathematical problems?

- Artistic Expression:

Arts questions: What is an artist able to express by creating or performing a piece of art? What is a visual artist able to express by creating or performing art?

Arts Integration question: How is expression in art about completing an art work similar to or different from what an author can express in language arts or a mathematician can express in problem solving tasks?

- Imagination in the Process of Arts Making:

Arts questions: What kind of ideas or images do you have when performing or making art? Where do you think these ideas come from? What kind of ideas or images do you have when making art? Where do you think these ideas come from?

Arts Integration question: How ideas and images in art also reflected in writing reading or math? How are these ideas and images similar to, or different from, ideas and images in art?

- Aesthetic Experience

Arts questions: What kind of responses or images do you have when attending carefully to a high quality piece of art? How can an experience of a piece of art possibly affect others? What kind of responses or images do you have when looking carefully at a high quality art? How can an experience of art possibly change someone’s understanding of him- or herself and/or his or her life?

Arts Integration question: What kinds of responses or images do you have when reading high quality books or appreciating math in solving problems? How are these ideas and images similar to, or different from, ideas and images in creating or experiencing art?

Transcriptions of the PAI interviews were then analyzed by IB-TAP researchers to determine levels of individual student response in both control and treatment schools based on the four-point scoring rubric below:

IB-TAP Student Response Rubric

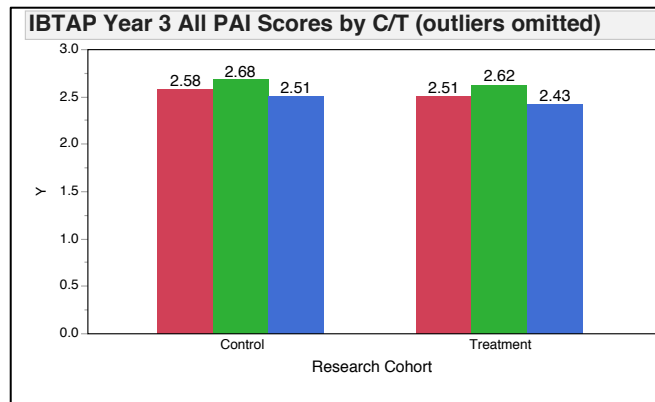
Categories of Response
Level 0: No Relevant Response; Irrelevant or indiscernible response; silence.
Level 1: Single Dimensional Responses

Concrete, un-detailed response. Generic statements, singular perspective. Unspecific, unfocused, diffused. No elaboration, no detail, no personal specifics or procedural relationships. Lists undifferentiated elements. “One dimensional thinking”
<p>Level 2: Multiple Single Dimensions</p> <p>Concrete connections, occasional detail, some elaboration, or emerging specificity; Some coordination of elements, like a clearly ordered procedure. Specific personal insight.</p>
<p>Level 3: Coordination of Dimensions of Understanding</p> <p>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.</p>
<p>Level 4: Systemic Understanding</p> <p>Substantial detail and specificity. Causal statements. Compare and contrast relationships. Critical perspective, highly complex, multiple relationships.</p>

Section 4B: IB-TAP Performance Assessment Interview (PAI) Results

After establishing a inter-rater reliability rating of 95% or above, the data were then entered and compared across control and treatment schools as indicated by Figure 15 below.

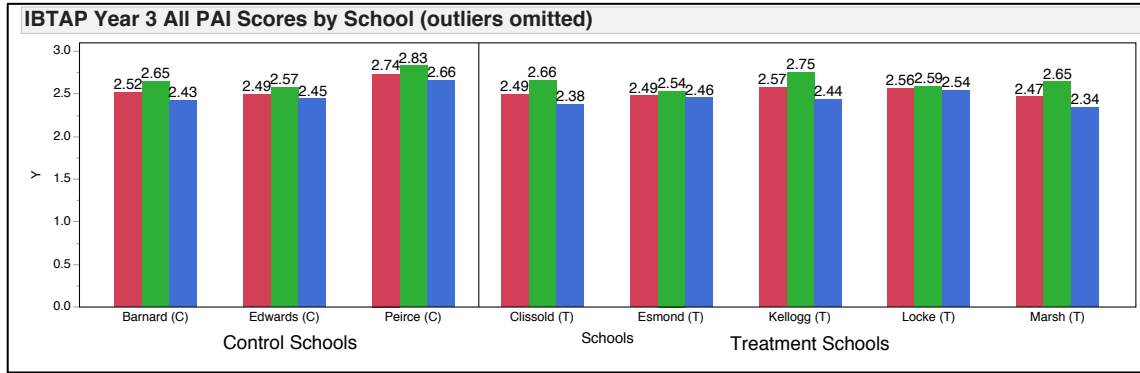
Figure 15: Control-Treatment School Overall Combined, Arts Integration, and Arts Domain PAI Rating Comparisons



The first column score represents the Overall average PAI score, the second column is the Arts Integration score, and the third column is the Arts Domain score. The PAI score data profiles for Control and Treatment are similar, differing from 0.04 – 0.08 in each category, with the Treatment

slightly behind in each. This finding confirms that, as a whole, the IB control and treatment schools both emphasize arts and arts integration equally.

Figure 16: C-T Individual School Distribution of PAI Averaged, Arts Integration and Arts Domain Scores



The individual school PAI scores have similar profiles. It appears also that the Peirce control school is largely responsible for the control schools having slightly higher ratings in each PAI category of response.

Section 4C: *The Relationship of Control and Treatment School Year 3 PAI ratings and Year 2 ISAT Combined Academic Scores*

The correlation between Year 3 PAI scores and Year 3 ISAT scores was not possible because of the irregularities of the ISAT transitional test invalidated the comparison as cited earlier in this report. Nonetheless, the correlation between Year 3 PAI and Year 2 ISAT was performed because of the longitudinal interpretation of the correlation data: that is, that the cumulative relationship between PAI scores can be seen across the years of the study.

However, the Table 2 correlation matrix reveals that there are no significant statistical links between indicators of levels of Arts/Arts Integration Interview Responses and Control School student academic achievement test scores.

Table 2: Matrix of CONTROL School PAI Correlations with ISAT Scores

	Year 3 PAI Combined Average Score	Year 3 PAI Arts Integration	Year 3 PAI Arts Domain
Year 2 ISAT Combined Avg	0.2603, p=0.1434	0.2947, p=0.1015	0.2714, p=0.1265

Score			
-------	--	--	--

^t = positive trend; * = significant (p value <.05); ** = very significant (p value < .01)

Table 3 below, however, shows that there is a significant positive trend between Treatment School PAI overall scores and a statistically significant correlation between PAI Arts Domain ratings with the Year 2 ISAT scores.

Table 3: Matrix of TREATMENT School PAI Correlations with ISAT Scores

	Year 3 PAI Combined Average Score	Year 3 PAI Arts Integration	Year 3 PAI Arts Domain
Year 2 ISAT Combined Avg Score	0.3511, p=0.0571 ^t	0.1902, p=0.3141	0.4056, p=0.0262*

^t = positive trend; * = significant (p value <.05); ** = very significant (p value < .01)

Section 4 Conclusions: Similarities in the relationship between PAI scores and ISAT reveals that both IB-TAP control and treatment schools emphasize arts and arts integration practices that are not necessarily associated strongly with standardized measures of academic achievement. However, in the Treatment Schools that enjoy the added presence of the teaching artist in academic classrooms, PAI scores in particular related to question about arts (as described in Section 4A) are significantly associated with ISAT scores, and therefore may be seen as significant contributor to the academic “baseline to Year 2” progress differences between the control and treatment schools reported earlier.

In the next section of the report, correlations between the PAI scores and the Student PC scores will be examined.

* * *

Section 5: IB-TAP Treatment School Student Portfolio Conference Ratings

Section 5A: IB-TAP Portfolio Conference Methods

The Teacher-Student Arts/Arts Integration Portfolio Conference (PC) Interview protocols were developed and adapted by Dr. Lawrence Scripp. The student section of the protocol was used for determining levels of response to questions about their classroom work with teaching artists, the work

of their peers, and to describe how and why their IB-TAP units contribute to their classroom learning and performance on standardized academic tests.

Students were required to bring their IB-TAP Art Integrated-Developmental Workbooks (AI-DW) and teachers attending the sessions provided contextual documents (relevant curriculum unit plans, class assignments, and evaluation rubrics) that were made available to interview facilitator prior to the start of the conference to look over the scope of work to be discussed that day.

Three students, one from each previously designated High, Average, and Low Achieving category were selected to participate in the first portions of each conference; the facilitator was not made aware of each student's classification. Each student was challenged to select one favorite work sample that illustrates well the most interesting things they did that year that they wanted to talk about during the session. The students individually and as a group then were challenged to demonstrate and reflect on their work for the IB-TAP projects in their school, describe the function of the teaching artist in the classroom, assess the relevance of arts integration teaching and learning in their classroom in relation to traditional IB classroom practices, and describe their best moments of learning during their IB-TAP units. They also were asked to participate in performance assessment tasks that require demonstrating their projects, guided debates with their peers, and assessing their work and the work of their classmates.

The Protocol lasted 45-50 minutes and ended with teachers reflecting on student performance saw during the conference.

Transcriptions of the Portfolio Conference events were made from video recordings and later analyzed by IB-TAP researchers to determine four levels of individual students levels of response in both control and treatment schools based on the rating rubric similar to the one used for the PAI interviews

Section 5B: *The Relationship of Treatment School Only Year 2 and Year 3 PC ratings and Year 2 ISAT Combined Academic Scores*

The correlation between Year 2 PAI scores and Year 3 ISAT scores was not possible because of the irregularities of the ISAT transitional test invalidated the comparison as cited earlier in this report. Nonetheless, the correlation between Year 3 PAI and Year 2 ISAT was performed because of the longitudinal interpretation of the correlation data: that is, that the cumulative relationship between PAI scores can be seen across the years of the study.

Section 5C: Relationships Between Student Portfolio Conference (PC) and Student PAI or ISAT Ratings

Table 4 below reveals that, by Year 3 of the project, Student PC ratings correlate positively with PAI total average score; the PC also correlates with PAI Arts Domain Ratings to a high degree of statistical significance.

Table 4: Matrix of Treatment School Student PC Correlations with PAI Scores

	Year 2 Student PC Ratings	Year 3 Student PC Ratings
PAI Year 3 Total Average	0.2323, p=0.2085	0.4169, p=0.0601 ^t
PAI Year 3 Arts Integration	0.1628, p=0.3816	0.3273, p=0.147
PAI Year 3 Arts Domain	0.2594, p=0.1588	0.4800, p=0.0276*

t = trend; * = significant (p value <.05); ** = very significant (p value < .01)

Table 5 below reveals that Year 2 and 3 Student PC ratings, though statistically insignificant, correlate relatively strongly with Year 2 ISAT scores.

Table 5: Matrix of Treatment School Student PC Correlations with Year 3 ISAT Scores

	Year 2 Student PC	Year 3 Student PC
Year 2 ISAT Comb Avg	0.1494, p=0.4665	0.3206, p=0.2260

t = trend; * = significant (p value <.05); ** = very significant (p value < .01)

Section 5 Conclusions: Positive correlations between Student Portfolio Conferences and Performance Assessment Interviews show that these two alternative assessments of arts/arts integration teaching and learning capture similar aspects of IB-TAP program student outcomes. While the PC ratings are not strongly directly related to ISAT scores, these PC data will be entered into stepwise regression analysis to determine its predictive value for academic achievement, controlling for student demographic values and teacher quality ratings.

* * *

Section 6: IB-TAP Treatment School Teacher Professional Development Outcome Variables

Section 6A: *Methods for investigating the question: Is there statistical evidence linking teacher professional development ratings to student learning outcomes in IB-TAP treatment schools?*

In order to investigate this question, the IB-TAP co-principal investigators examined teacher professional development outcome variables that would systematically measure the classroom teachers' application of IB-TAP program concepts to their classroom practices. The instruments used for rating the teachers according to their quality of participation in the project included:

- The IB-TAP Effective Teaching and Student Engagement Observation Protocol. This tool consisted of items designed to record and analyze two lessons (beginning and end) in each treatment teachers' classroom over the course of their IB-TAP Arts Integrated Interdisciplinary Units. The constructs that enabled a consistent look of creative teaching across IB-TAP classrooms included the IB-TAP Creative Cycle (Page 4) and additional items adapted from the Center for Research on Education, Diversity and Excellence's (CREDE) Observation Protocol. This protocol highlights effective research-based teaching practices found to facilitate student engagement, including collaborative teaching and learning, working together, connecting instruction to students' lives, and providing challenging activities.
- Teacher Portfolio Artifact Analysis Tool with Rubric and Scoring Guide. A rubric and scoring guide were developed in order to analyze curriculum artifacts such as the mid year curriculum planning documents, and the Teacher Portfolios.
- The IB-TAP Teacher-Student Arts Integration Performance Assessment Portfolio Conference Protocol was designed by co-principal investigator Lawrence Scripp and analyzed by his research team in order to capture evidence of teacher professional development and student learning outcomes. This 45 minute performance assessment and interview protocol provided opportunities for (a) teachers to describe the goals and practices of the program, (b) students to present, elaborate, and demonstrate their understanding of the IB-TAP project and portfolio work and its connection to math, social studies and English language arts, and (c) teachers to reflect on the performance of their students during the conference. Video analysis of the entire portfolio conference was guided by rating scales designed to capture evidence of categories and

degrees of reflective understanding of the arts integration concepts and practices documented in student IB-TAP portfolios.

Further description of these tools and how they were employed to rate teacher effectiveness is presented in Louanne Smolin's Co-Principal Investigator's Evaluation Report. Her report also includes statistical analysis of the degree of improvement in the teacher practices most likely to optimize arts integration teaching and learning practices in IB-TAP. The teachers' reflections on the project suggested that arts-integrated curriculum and collaboration with teaching artists had impacted their learning and teaching practices.

The process of teacher ranking variables derived from Teacher Interview transcripts, IB-TAP Portfolio Conference Teacher interview transcripts, IB-TAP Effective Teaching Observational Data, and indication of project participation, such as Years in Project and Number of Professional Development Events attended. The hallmark of teacher learning was indicated by (a) increasing effective collaborative practices, (b) understanding of the creative process and its application to their arts integration teaching practices, (c) documenting their curriculum by gathering a greater variety of artifacts that better represent the quality of student work, and (d) their efforts to self reflect and assess their own teaching in their documentation in process journals and reflection writing.

From all of these efforts, the following independent teacher rating variables related to aspects of the IB-TAP Program included the following:

- Collaboration between IB-Teachers and External Arts Partner Teaching Artists.
- Arts Integration through Collaboration
- Documentation Practices that consisted of multimedia artifacts to create narratives of the curriculum and capture evidence of teacher and student learning.
- Reflection on teaching and learning in order to refine curriculum, note teacher professional learning, and better understand students as learners.

* * *

Section 6A: Interrelationships between Teacher Professional Development Variables and Student Learning Outcomes

As Table 6 makes clear, despite all of the statistical evidence for teacher improvement throughout the evolution of the IB-TAP project, there is no correlational evidence of any link with student learning outcomes.

Table 6: Year 2 Correlation Matrix Examining the Relationship Between Teacher Rating Variables and Student Learning Outcomes

	Year 2 Teacher # of Years in Program	Year 2 Teacher # of PDs attended	Year 2 Adjusted Teacher Rating	Cumulative Teacher Adjusted Rating	Year 2 PC Teacher Score
Year 2 ISAT Comb Avg	0.0256, p=0.9032	0.0208, p=0.9214	0.1545, p=0.4816	-0.0042, p=0.9810	-0.3171, p=0.1145
Year 2 Student PC Score	0.1436, p=0.4035	-0.1505, p=0.3810	-0.0409, p=0.8211	-0.0495, p=0.7845	-0.0995, p=0.5468
Year 3 PAI Avg Score	-0.0588, p=0.7532	-0.0002, p=0.9991	-0.0610, p=0.7577	0.0632, p=0.7227	0.0628, p=0.7373
Year 3 PAI Arts Integration	-0.1079, p=0.5634	-0.1691, p=0.3631	0.0900, p=0.6486	0.1770, p=0.3167	0.0619, p=0.7410
Year 3 PAI Arts Domain	0.0294, p=0.8751	0.2027, p=0.2741	-0.2104, p=0.2824	-0.0682, p=0.7017	0.0864, p=0.6441

Section 6 Conclusions: The lack of evidence for any link between teacher and student learning—even at the basic level of PD attendance or years in the project—suggests that IB-TAP classroom teacher practices had little to no impact on arts or academic learning. This claim will be tested further by entering teacher quality and participation variables into the stepwise regression equation.

Besides inadequate measures of the teacher evaluation, another possibility is that the quality of the teaching artists and individual student work in portfolios may be missing ingredients in this analysis. Without collecting data on teaching artist quality or the quality of teaching artist-classroom collaboration, or differences in arts integration portfolio work, the evaluation process may be missing the best predictors of student learning outcomes.

* * *

Section 7: Results from Stepwise Regression Multivariate Analysis

Stepwise regression techniques were administered in order to test the influence of all student learning outcomes, teacher factors, and student demographics on the Year 2 ISAT academic scores. As indicated by Figure 16 the final regression equation accounted for 99.7 percent of the variance (R^2) in the ISAT scores.

In this particular case, one student learning outcome—Student PC Ratings (Estimate 81.84, $p=.02$) (5th line from the bottom)— and one teacher variable—Years in the Project (Estimate 04.01 $p=.05$) (2nd line from the bottom)—exerted the most influence on ISAT scores.

Though Student PC wasn't the only factor that predicts ISAT, yet in interaction with the other factors Student PC performance ratings were able to positively predict ISAT more than any other variable, including all student demographics and teacher variables. It was the second factor to be added into regression, moving R^2 from 0.6494 to 0.7128 (1112 ISAT with Teacher Adjusted Rating), which is not a significant increase in amount of prediction at that point in the model. Yet its significance in the overall model is probably due to the fact that Student PC Ratings are linked to so many of the other measures matters more than the absolute importance of it to the model.

Figure 16A Stepwise Regression Model Targeting Year 2 ISAT Student Outcome

Stepwise Fit for 1112 ISAT Combined-Average Scores (outliers omitted)									
SSE	DFE	RMSE	RSquare	RSquare Adj	Cp	p	AICc	BIC	
27.231805	23	3.6899732	0.9971	0.9757	16	16	.107	.6704	
Current Estimates									
Lock	Entered	Parameter	Estimate	nDF	SS	"F Ratio"	"Prob>F"		
		Intercept	715.486728	1	0	0.000	1		
		Ethnicity{Multi&Afro-Am&Hispanic-Asian}	16.6604834	3	1091.734	26.727	0.03628		
		Ethnicity{Multi&Afro-Am-Hispanic}	38.0913201	2	422.9947	15.533	0.06048		
		Ethnicity{Multi-Afro-Am}	-30.218699	1	80.7428	5.930	0.13525		
		Gender{Male-Female}	-14.845458	1	135.4923	9.951	0.0875		
		Free Lunch Program?{Yes-No}	-16.92231	1	214.0269	15.719	0.05813		
		Student HAL Baseline Academic Designation{L&A-B}	17.786441	2	749.8554	27.536	0.03504		
		Student HAL Baseline Academic Designation{L-A}	32.5130958	1	306.6753	22.523	0.04164		
		1213 PAI Total Avg of all PAI scores	-4718.3621	1	142.8742	10.493	0.08353		
		1213 PAI Avg of arts integration scores	1789.40132	1	140.4893	10.318	0.08478		
		1213 PAI Avg of Arts Domain scores	2886.6368	1	142.7019	10.481	0.08362		
		1112 Student PC Score	81.8404833	1	784.9399	57.649	0.01691		
		1112 Teacher Rating	-41.21036	1	143.5748	10.545	0.08317		
		1112 Teacher PC Score Question Average	-190.90879	1	211.8429	15.558	0.05867		
		1112 Teacher Number of Years in Program	94.9112781	1	229.3752	16.846	0.05455		
		1112 Teacher Total # of PD sessions attended	-19.226475	1	190.5402	13.994	0.06461		

Figure 17B: A Record of the Variable Entry Sequence

Stepwise Fit for 1112 ISAT Combined-Average Scores (outliers omitted)										
SSE	DFE	RMSE	RSquare	RSquare Adj	Cp	p	AICc	BIC		
27.231805		23.6899732	0.9971	0.9757	16	16	.107.6704			
Step History										
Step	Parameter	Action	"Sig Prob"	Seq SS	RSquare	Cp	p	AICc	BIC	
1	Student HAL Baseline Academic Designation{L&A-E}	Entered	0.0001	6198.24	0.6494	231.8	2	152.853	153.81	
2	1112 Student PC Score	Entered	0.0887	605.6602	0.7128	189.32	3	152.622	153.107	
3	Free Lunch Program?{Yes-No}	Entered	0.1345	418.5522	0.7567	160.58	4	153.563	153.015	
4	Gender{Male-Female}	Entered	0.1749	317.5645	0.7899	139.25	5	155.552	153.258	
5	Ethnicity{Multi&Afro-Am-Hispanic}	Entered	0.1326	616.4404	0.8545	97.979	7	161.303	152.426	
6	Student HAL Baseline Academic Designation{L-A}	Entered	0.3994	99.8539	0.8650	92.646	8	168.46	153.973	
7	1112 Teacher Number of Years in Program	Entered	0.2007	225.2316	0.8886	78.104	9	175.931	153.406	
8	1112 Teacher Total # of PD sessions attended	Entered	0.0223	531.2455	0.9442	41.087	10	178.042	143.836	
9	1112 Teacher PC Score Question Average	Entered	0.0406	251.728	0.9706	24.6	11	186.912	135.197	
10	1213 PAI Avg of arts integration scores	Entered	0.2345	63.23787	0.9772	21.955	12	212.914	133.488	
11	1112 Teacher Rating	Entered	0.3124	43.75999	0.9818	20.741	13	259.865	132.33	
12	Ethnicity{Multi-Afro-Am}	Entered	0.7992	3.146538	0.9822	22.51	14	361.535	134.891	
13	1213 PAI Total Avg of all PAI scores	Entered	0.9379	0.405346	0.9822	24.481	15	667.492	137.738	
14	1213 PAI Avg of Arts Domain scores	Entered	0.0836	142.7019	0.9971	16	16	.107.67		

Section 7 Conclusion: The final regression analysis reinforces the fact that despite flaws and unintended anomalies in the project design, student arts learning is linked strongly with academic performance in the IB-TAP project. That the Arts/ Art Integration Student Portfolio Conference Rating are more predictive of ISAT performance than are factors such as ethnicity, social economic status, and previous academic performance is impressive; the fact that teacher variables cannot be linked with academic results or arts/arts integrated learning assessments is cause for concern. Arts researchers and program directors should be looking deeper into ways to evaluate teacher quality in arts integration projects and for tools measure more effectively the influence of student work and teacher artist input into the arts integration policy equation.

* * *