National Museum of Women in the Arts
Art, Books, and Creativity

Final Report

2007

Prepared by:

Dr. Robert W. Lissitz
Dr. Melissa Fein

TABLE OF CONTENTS

TABLE OF TABLES

Introduction

This report is the final program evaluation document covering the entire duration of the *Art, Books, and Creativity* program implementation. *Art, Books, and Creativity* is a program designed by the National Museum of Women in the Arts (NMWA) to integrate visual arts and language arts into classroom lessons for elementary school students. The primary goals of the project were to promote the acquisition of basic skills in creating and responding to the visual arts; to further an interdisciplinary method of learning through the arts; to expand the creativity, critical thinking, and communication skills of the students; to transform the American school arts curriculum into a model that includes information on women artists; and to create a curriculum that can be widely replicated. *Art, Books, and Creativity* project partners include the following individuals and organizations: independent evaluator Dr. Robert Lissitz, professor and former chair of the Department of Measurement, Statistics, and Evaluation at the University of Maryland; evaluation consultant Dr. Melissa Fein of *Measuring Merit*; Albuquerque Public Schools, Albuquerque, N.Mex.; the Albuquerque Museum of Art and History, Albuquerque, N.Mex.; and Arlington Public Schools, Arlington, Va.

The evaluation plan for the *Art, Books, and Creativity* program was designed to determine the extent to which program goals have been achieved by utilizing the following multi-step process:

1. Outlining operational objectives that correspond to overriding program goals;
2. Reframing these objectives as evaluation questions;
3. Identifying measurable outcomes that can be used to answer these questions;
4. Finding or developing the appropriate instruments to use as measurement tools; and
5. Collecting, analyzing, and interpreting the data.

The overriding program goal was the improvement of visual literacy skills. The evaluation question of primary interest was whether students participating in *Art, Books, and Creativity* had greater gains in visual literacy and written expression related to art concepts than those students who were not in the program. The *Art, Books, and Creativity* Visual Arts Assessment was developed specifically for this project as a measure of fourth grade visual literacy and written expression related to art concepts. Although random sampling was not possible, efforts were made to balance the experimental and control schools by key demographics. The *Art, Books, and Creativity* Visual Arts Assessment was comprised of a pre-test and a post-test. To determine the effectiveness of the *Art, Books, and Creativity* program, the evaluators used Hierarchical Linear Modeling (HLM) as the statistical approach to analyze the gains in visual literacy achievement data. Focus groups and teacher surveys were used to add a qualitative dimension to the visual arts achievement results as well as to evaluate process and implementation aspects of the project. This evaluation report includes an outline of the program objectives, an outline of the evaluation questions corresponding to those objectives, an overview of the evaluation

design and methodology, an outline of the sample selection procedures, a description of the data collection process, and a summary of the overall program results over both academic years during which the program was implemented.

Program Objectives

The main components of *Art, Books, and Creativity* were lessons that introduce students to basic art concepts and vocabulary, journal writing, a museum visit, workshops with visiting artists, and the creation of artists' books. Through the implementation of these components *Art, Books, and Creativity* was oriented toward the achievement of the following objectives:

1. Increasing and enhancing student knowledge of visual literacy concepts and vocabulary;
2. Increasing and enhancing student written expression related to art concepts;
3. Increasing and enhancing student artistic self-confidence;
4. Supporting and enhancing the existing district art education objectives; and
5. Providing teachers with high quality training workshops designed to enable teachers to support objectives one through four above.

Evaluation Questions

The evaluation of *Art, Books, and Creativity* addressed and measured the extent to which the program achieved its major goals and objectives by answering the following questions:

1. Did students who participated in *Art, Books, and Creativity* have greater gains in visual literacy concepts and vocabulary than those who were not program participants?

2. Did students who participated in *Art, Books, and Creativity* show greater gains in reflective written expression related to art concepts than students who were not program participants?

3. Did students who participated in *Art, Books, and Creativity* show greater gains in artistic self-confidence than students who were not program participants?

4. Did *Art, Books, and Creativity* support and enhance the existing district art education objectives?

5. Were *Art, Books, and Creativity* professional development workshops appropriate and effective?

Evaluation Design

The evaluation design for *Art, Books, and Creativity* included the integration of qualitative as well as quantitative data. The integration of qualitative and quantitative information follows levels one and two of the Kirkpatrick model of evaluation (Kirkpatrick, Donald L. *Evaluating Training Programs: The Four Levels*, 2nd ed., San Francisco: Barrett-Koehler, 1998). Quantitative data was used to assess the achievement gains in visual literacy by comparing pre-test and post-test scores, and was used to address evaluation questions one, two, and three. Interviews and focus groups with teachers were used to answer evaluation questions two though five, and thus, this provided multiple perspectives on evaluation questions two and three.

Teacher survey data from the teachers' survey in the spring of 2005 were used as a formative tool in the sense that feedback was used to make program improvements for the 2005–2006 program year. The Year Two teacher surveys addressed the extent to which these program improvements were effective. This evaluation report presents an analysis of the overall *Art, Books, and Creativity* program results over the entire duration of the program.

Methodology

The purpose of this methodology section of the evaluation report is to provide the details of how student achievement gains were measured, analyzed, and interpreted. The section on instrument development includes an outline of the table of specifications, a description of the piloting process, a description of the rubric used to grade the written expression component of the instrument, scoring details for the multiple-choice component of the exam, and administrative details. The rationale for the quantitative method of analysis that was selected is provided. In addition, the rationale for the collection and use of qualitative data is outlined. Also covered are the sample selection process, demographics of the sample, and the general data collection plan and retention issues.

Instrument Development: Measuring Visual Literacy

It was hoped that an existing visual literacy instrument with known reliability could be used to assess achievement gains, but no suitable instrument was located. In January and February 2004, the project evaluators, Dr. Fein and Dr. Lissitz, worked with NMWA project staff to design a visual arts concepts instrument specifically for *Art, Books, and Creativity*. A draft of the instrument was completed in March 2004. Drafts were reviewed by school officials for age appropriateness, as well as for general appropriateness and adequacy.

Table of Specifications

The instrument used to evaluate the effectiveness of *Art, Books, and Creativity* was intended to measure students' visual literacy skills and, because the program integrates visual arts and language arts, the instrument also includes a written expression component. The corresponding learning objectives were that students would be able to define categories of subject matter such as portrait, landscape, narrative, and still life; recognize basic elements of art (color, line, shape, form, pattern, and texture); understand the artistic process of abstraction; and apply an understanding of these terms and concepts to the discussion and creation of artworks. It should be noted that the test objectives were designed to reflect the *Art, Books, and Creativity* program goals, as well as curriculum goals common to national and district (Arlington, Va., and Albuquerque, N.Mex.) art learning objectives.

The tests (parallel pre-test and post-test) comprised twenty multiple-choice items and one essay item. The multiple-choice items focused on measuring visual literacy. The short-answer essay question was designed to assess student ability to interpret a work of art using correct art vocabulary and expressive language. Two additional items were embedded in the multiple-choice test and were designed to measure student confidence in creating and talking about art. The responses to these two items are not included in the achievement scores; they were used to track the change in affective response over time.

The table of specifications covering the number of items for each topic area is presented in Table 1.

Table 1. Table of Specifications

Topic	Number of Items
Overarching Concepts	2
Categories of Subject Matter	2
Interpretation	4
Line	1
Color	4
Perspective	2
Pattern and Rhythm	2
Form and Shape	2
Abstraction	1
Total	**20**

Instrument Piloting

In June 2004, the *Art, Books, and Creativity* evaluation instruments were piloted with two fourth grade classes in Arlington; forty-five students were tested. The completed pilot tests were forwarded to project evaluator Dr. Fein for scoring and item analysis. The results of the pilot were used to estimate the instrument reliability and to identify items that needed revising. The reliability of the pre-test pilot as measured by coefficient alpha was 0.75. Item analysis and teacher feedback indicated that two items needed to be revised for the final version. The pre-test revisions were completed in July 2004. The initial reliability of the post-test pilot as measured by coefficient alpha was 0.54. The post-test required more revisions than the pre-test and the revised version was re-piloted before its use in the spring of 2005. The coefficient alpha for the revised version was 0.59. The reliability measure, coefficient alpha, varied some according to the sample on which it was computed. For the pre-test in Year One, the coefficient alpha was 0.62, and in Year Two it was 0.64. The coefficient alpha for the post-test in Year One was 0.71 and for Year Two it was 0.72.

Rubric

The rubric for the essay questions was developed in conjunction with the development of the essay question to ensure a match between the learning objective and this component of the instrument, and it was further refined after the pilot. A copy of the rubric is provided in the appendices.

Instrument Administration

The pre-test was administered to all experimental and control classes in the fall of the 2004–2005 and the 2005–2006 school years. Matched pre-tests and post-tests for the multiple-choice component, the essay component, and the composite were collected from 1,264 students. More students were tested, but due to mobility and absentee issues, all scores were not available for all students. The same instrument was not used for the pre-

test and the post-test because this would pose a "test-retest" threat to validity, therefore the pre-test and post-test were designed to be parallel tests.

The pre-tests and post-tests were administered by the classroom teachers. The teachers were given standardized instructions for test administration. After the pre-test in Year One, the instructions were revised based on evaluation as well as informal feedback.

Instrument Scoring

The multiple-choice portion of the answer sheets was coded on NCS Pearson General Purpose Answer Sheet Form Number 4521. The completed tests were forwarded to the evaluator, Dr. Fein, who sent the answer sheet forms to Infoscan in Bethesda, Md., for scanning. The essay items were scored using the rubric. The grader of the essay items, Nancy Jakubowski, was trained by Dr. Fein in the use of the rubric. Jakubowski holds a bachelor's degree in English Literature and a master's degree in education and human development with a major concentration in adult education, both from George Washington University. Jakubowski has extensive professional, as well as volunteer experience in the field of education. She graded the pre-tests and post-tests in Year One and Year Two. The use of the same grader removed the concern of inter-rater reliability between pre-test and post-test results, and assured consistency of grading from year to year.

One of the key components of *Art, Books, and Creativity* was the integration of reflective writing and journal writing in the creation of artwork. It was not financially feasible to evaluate the artwork and associated writing for all students. Even if a sampling of the projects were graded, it would have been impossible to compare the experimental and control groups since the control group did not create artists' books. Information from both the focus groups and the teacher surveys were used to provide evaluation information on this key component of the program.

Analyzing Achievement Data

The focus of the *Art, Books, and Creativity* program data analysis was to compare the growth in achievement scores of the program participants to the growth in achievement scores of non-participants (the control group). Growth in achievement was measured by the difference between the post-test and pre-test scores. Two areas of achievement were measured: visual literacy and written expression related to visual arts concepts. Visual literacy was measured through the use of a multiple-choice component of the instrument. Written expression related to art concepts was measured through the use of an essay component of the instrument. The combined score of these two components of the instrument is referred to as the composite score. Achievement growth or gains were analyzed for the multiple-choice component, the essay component, and the composite. Conventionally, analysis of variance (ANOVA) is used to compare the data of program participants and the control group. However, when data are nested (in the case of students nested within classes, which are nested within schools), inferential validity of the ANOVA estimates can be compromised by threats of misestimated precision (underestimated standard errors) (Raudenbush and Bryk, 1988, p. 429). Hierarchical modeling is an analytic technique developed as a response to shortcomings in

traditional/conventional statistical approaches that occur when applied to multilevel (nested) data. The *Art, Books, and Creativity* project demographics were quite different in the two regions, and the program was administered slightly differently in each regional location; in Albuquerque the program was taught by classroom teachers and in Arlington it was taught by art teachers. The *Art, Books, and Creativity* project data have a hierarchical structure, and hierarchical linear modeling the most appropriate method of data analysis to use to compare the achievement growth of program participants to the achievement growth of the control group. The analysis in this report is based on all pre-tests and post-tests collected during both years that the program was implemented.

Analyzing Survey Data and Qualitative Responses

Data from the teacher surveys were tabulated. In addition, comments from the interviewers and focus groups were used to explain and enhance the survey results and the student achievement data. In this report, only key overall survey and focus group results are reported based on data from both years of the program. The complete survey and focus group reports for each year of the program are located in the appendix of the annual grant reports.

Development of the Teacher Survey

Teacher surveys were developed from focus groups held at the close of the academic year at each site. One teacher was selected randomly from every school for focus group participation. The teacher survey results and comments from the teacher interviews and focus groups were used to assess the achievement of evaluation questions two through five: the extent to which students showed gains in reflective expression; the extent to which students showed gains in artistic self-confidence; the extent to which *Art, Books, and Creativity* supported and enhanced existing district arts education learning objectives; and the extent to which professional development workshops were appropriate and effective.

Sample Selection

Experimental Group

Art, Books, and Creativity was specifically designed to serve the needs of students at risk, and the schools selected to participate in the program reflected this priority. In February 2004, arts education specialists in the Albuquerque and Arlington public school districts began the selection process through an open application invitation. In Arlington the process was competitive—art teachers were required to submit applications in which they discussed how the project would impact them professionally and how their students and their schools would be affected. Arts Education Specialist Pam Farrell reviewed the applications and made the final selections with evaluators Dr. Fein and Dr. Lissitz, who compared student demographics to encompass a broad range of test scores, ethnicity, socioeconomic levels, and English Language Learners (ELL). In Albuquerque, school principals were recruited for the program by Fine Arts Instructional Manager Janet Kahn. Again, the final selection of schools was made in conjunction with the evaluators to find comparable student demographics. Principals at the schools selected the fourth grade

teachers for participation. Over the two years that the program was implemented, there were twenty-five schools and forty-six teachers who participated in the program.

Control Group

The control schools were matched as closely as possible to the experimental schools by considering criteria such as percentage of students receiving free and reduced-price meals (FARMS), percentage of ELL, and language arts pass rates. Consideration of these criteria resulted in the control group being as similar as possible to the experimental group. The language arts pass rates for the two sites were based on two different instruments, and are not comparable across sites for that reason. The main measure that is comparable across sites for both years of the program is FARMS. A comparison of average FARMS for the experimental and control schools at each site for each year are presented in Table 2. The tables of the complete demographic profiles of the experimental and control schools used in the analysis of program effects are provided in the appendices. The rates for the two sites differ, but within each site for each year, the average FARMS rate for the experimental and control groups was not significantly different.

Table 2. FARMS Status: Group Averages Overall, by Site and by Year

Year and Site	Experimental	Control
Year One Albuquerque	81%	85%
Year Two Albuquerque	83%	79%
Year One Arlington	48%	51%
Year Two Arlington	51%	49%
Overall Albuquerque	**82%**	**82%**
Overall Arlington	**50%**	**50%**

It was of paramount importance that *Art, Books, and Creativity* serve the student populations of Title 1 schools. There were no significant differences in the average socioeconomic status (as measured by FARMS) between the experimental schools and control schools. To motivate schools to be in the control group, comparable schools not selected for participation in the first year were given first priority of selection to participate as experimental schools during the second year. This explains why some control schools in Year One reappear as experimental schools in Year Two.

Data Collection

Pre-tests and post-tests assessing visual literacy skills were administered to the participating students at the beginning and at the end of the *Art, Books, and Creativity* program. The data collection schedule for Year One was to administer pre-tests in September 2004, before program implementation; to administer post-tests in May 2005; hold teacher focus groups in May 2005, and administer teacher surveys in June 2005. In Year Two, the administration of the post-tests was scheduled earlier in the year so that it would not overlap with annual state testing. This change was made in response to teacher

input from the Year One focus groups. The data collection schedule for Year Two was to administer pre-tests in September 2005 and post-tests in March and April 2006; hold teacher focus groups in May 2006; and administer teacher surveys in June 2006.

Retention

For both academic years of the program, 2004–2005 and 2005–2006, fourth grade students were the focus of the study. This program was embedded in the school program, with buy-in from the arts education personnel and administrators. In Arlington the program was taught by art teachers during their once-per-week session with students. In Albuquerque the program was taught by classroom teachers during their regular class time. Not only were the locations demographically different, but the programs were administered differently in each location.

It should be noted that in each year of the program that the participants were fourth grade students. The same group of students were not followed; the program was not concerned about retention from one program year to the other. The main retention problems reflect the general problem of student mobility during a given school year, and truancy. Another retention issue was ensuring that the teachers who were trained to implement the program before the start of the school year would actually be teaching at the participating schools once the school year began.

Results

The results summarize the following:

- Achievement gains;
- Gains in artistic self-concept/self-confidence;
- The extent to which the *Art, Books, and Creativity* program supported district visual arts learning objectives;
- The quality of the professional development workshops; and
- The extent to which process and implementation changes made in response to Year One evaluation findings were effective.

Achievement Gains

Visual Arts Concepts and Vocabulary: Multiple-choice Component

To evaluate whether *Art, Books, and Creativity* participants had greater gains than the control group in the acquisition and use of visual arts concepts and vocabulary, "difference scores" for the two groups were compared. The difference scores are defined as the difference between the pre-test and post-test scores. These differences provide a measure of the gain in student knowledge using data from both years of the program. Sometimes these difference scores are referred to as gains or gain scores. To measure achievement gains in visual arts concepts and vocabulary, the gain scores of the multiple-choice test scores were analyzed. Multilevel/hierarchical linear models were estimated to determine if the gain scores for the experimental and control groups were significantly different. Average score gains on the assessment components (multiple-choice and essay) and the composite score are presented in Table 3.

Table 3. Experimental and Control Group Score Gains on the Assessments

Group		Gain: Multiple-choice[1]	Gain: Essay[2]	Gain: Composite[3]
Control	Mean	0.96	0.37	1.33
	N	530	530	530
	Std. Deviation	3.32	0.96	3.60
Experimental	Mean	3.08	0.61	3.69
	N	734	734	734
	Std. Deviation	3.58	0.81	3.77
Total	Mean	2.19	0.51	2.70
	N	1264	1264	1264
	Std. Deviation	3.63	0.88	3.88

[1] There were twenty possible total points on the multiple-choice component of the assessments.
[2] There were four possible points on the essay component of the assessment.
[3] There were twenty-four possible points on the composite of the multiple-choice and essay assessments.

The average *Art, Books, and Creativity* participant score gain on the multiple-choice component of the assessment was significantly higher than the average control group score gain. The average *Art, Books, and Creativity* participant score gain on the multiple-choice test was 3.07 raw score points (there were a maximum of twenty points on the test), and the average control group score gain was 0.96 raw score points. The multiple-choice score gain was based on 1,264 matched pre-test and post-test scores collected during the two years that the program was implemented.

The statistical models, parameter estimates, and significance tests pertaining to these results are presented in detail in the appendices along with a breakdown of pre-test and post-test scores on the multiple-choice component for the experimental and control groups.

Written Expression/Reflective Writing: Essay Component

The average *Art, Books, and Creativity* participant score gain on the essay component of the assessment was significantly higher than the average control group score gain over the two years that the program was implemented. In Year One the gains in the written expression component were not significantly greater for participants, and it was surmised that encouraging fuller use of the writing component would remedy this. During the second year of the program, the writing component was more fully implemented, and the quantitative analysis of Year Two data indicates that written gains were significantly greater for the participants. **In addition, the program effect during Year Two was greater for students in schools with higher FARMS rates (schools with more students of lower socioeconomic status)**. The average *Art, Books, and Creativity* participant score gain on the essay item was 0.61 raw score points (there were a maximum of four points on the essay), and the control group score gain was 0.37 raw score points.

These results agree with teacher comments made during the focus groups. In their subjective assessment, teachers in the Arlington schools with higher FARMS rates, did not note improvements, whereas Albuquerque teachers in schools with higher FARMS rates did believe that the program had a positive effect on student writing.

The statistical models, parameter estimates, and significance tests pertaining to these results are presented and discussed in more detail in the appendices along with a breakdown of pre-test and post-test scores on the essay for the experimental and control groups.

Overall Achievement Gains: Composite Score

The average *Art, Books, and Creativity* participant score gain on the composite score (composite of essay and multiple-choice components) was significantly higher than the average of the control group score gain over the two years that the program was implemented. The average *Art, Books, and Creativity* participant score gain on the composite was 3.7 raw score points (there were a maximum of twenty-four total raw score points), and the control group score gain on the composite was 1.3 raw score points.

The statistical models, parameter estimates, and significance tests pertaining to these results are presented and discussed in more detail in the appendices along with a breakdown of pre-test and post-test scores on the composite for the experimental and control groups.

Affective Changes: Artistic Self-Confidence

Two attitude questions were embedded into the visual arts component of the pre-test. It was recognized from the outset that the two questions on attitude did not provide a reliable measure of affect. It was decided that teachers could not be overburdened with assessments beyond the pre-test and post-test, and that the two items embedded in the achievement test would not burden the teachers, yet would hopefully give some insight into changes in attitude related to artistic self-concept/self-confidence. These attitude questions gave a general idea of the attitudes held at two points in time by the experimental and control group students on these specific questions. Although these attitude items were derived from an attitude instrument of known reliability, care was taken in interpreting these results, since the use of only two attitude questions does not provide a reliable assessment of attitude change over time.

The first attitude question posed to the students was "How do you feel about your ability to make art?" The pre-test and post-test response data to this question for the control groups and experimental groups in both years and overall are summarized in Table 4. Overall, 86.9 percent of the control group students and 86.0 percent of the experimental group students indicated on the pre-test that they feel confident about their ability to make art. On the post-test overall, 84.3 percent of the students in the control group and 88.0 percent of the students in the experimental group indicated that they feel confident in their ability to make art.

Table 4. Percentage of students who indicated they feel confident about their ability to make art.[1]

Year and Group	Pre-test % of students (sample size[2])	Post-test % of students (sample size)
Year One Control Group	87.9% (323)	82.4% (323)
Year One Experimental Group	84.7% (426)	88.9% (431)
Year Two Control Group	86.0% (394)	85.8% (395)
Year Two Experimental Group	87.1% (504)	87.2% (508)
Overall Control Group	**86.9% (717)**	**84.3% (718)**
Overall Experimental Group	**86.0% (930)**	**88.0% (939)**
Overall	86.4% (1647)	86.4% (1657)

[1] Includes all students who indicated that they feel either "very confident" or "somewhat confident" and excludes those who indicate that they feel "not very confident" or "not at all confident."
[2] More attitude data were available than matched pre-test and post-test exam data, so these sample sizes will not necessarily match the sample sizes for the score gain analysis.

The second attitude question posed to students was "How do you feel about your ability to understand and talk about art?" The pre-test and post-test response data to this question for the control groups and experimental groups in both years and overall are summarized in Table 5.

Overall, 79.3 percent of the control group students and 78.1 percent of the experimental group students indicated on the pre-test that they feel confident about their ability to make art. On the post-test overall, 74.8 percent of the students in the control group and 77.8 percent of the students in the experimental group indicated that they feel confident in their ability to make art.

Table 5. Percentage of students who indicated they feel confident about their ability to understand and talk about art.[1]

Year and Group	Pre-test % of students (sample size[2])	Post-test % of students (sample size)
Year One Control Group	81.1% (323)	71.5% (323)
Year One Experimental Group	78.3% (423)	78.0% (432)
Year Two Control Group	77.8% (392)	77.6% (393)
Year Two Experimental Group	78.0% (499)	77.6% (503)
Overall Control Group	**79.3% (715)**	**74.8% (716)**
Overall Experimental Group	**78.1% (922)**	**77.8% (935)**
Overall	78.6% (1637)	76.5% (1651)

[1]Includes all students who indicated that they feel either "very confident" or "somewhat confident" and excludes those who indicate that they feel "not very confident" or "not at all confident."
[2] More attitude data were available than matched pre-test and post-test exam data, so these sample sizes will not necessarily match the sample sizes for the score gain analysis.

Enhancement of District Objectives

In order to answer the research question: "Does *Art, Books, and Creativity* support and enhance the existing district arts education learning objectives?" teacher surveys and focus groups responses were analyzed. Copies of the complete focus group reports and teacher surveys were submitted as appendices in the required annual grant reports and are not reproduced within this report.

In the Albuquerque focus groups, teachers stressed that the main strength of the program was its potential for integrating visual arts and language arts learning objectives. Teachers indicated that the student copies of *Exploring Art* and the reproductions of artworks available on CD or as overhead transparencies were very helpful for teaching the concepts and vocabulary needed to describe and analyze a work of art. Teachers emphasized that analyzing a work of art is similar to analyzing a piece of writing. In this respect *Art, Books, and Creativity* supported their schools' reading and writing programs,

particularly the writing of more detailed descriptions. Some teachers felt that many of their students were weak in the area of general knowledge, and that the program helped broaden the students' general knowledge base. In addition, teachers thought that the *Art, Books, and Creativity* curriculum supported and overlapped with the 6+1 Trait Writing framework that many of them use. Teachers emphasized that the *Art, Books, and Creativity* curriculum motivated the students to write because the students enjoyed the activities they were doing and took the time to do a really good job. When identifying what might be missing from the program, teachers indicated that lessons about sculpture were missing from the curriculum but were represented in the *Exploring Art* text.

In the Arlington focus groups, teachers said that the *Art, Books, and Creativity* curriculum aligned well with the fourth grade curriculum in their district. Although the teachers agreed that nothing key was missing from the curriculum, they also said that they supplemented the curriculum.

In Arlington, some art teachers indicated that classroom teachers were supportive of the writing component and others said they received no support from the classroom teachers. Some Arlington art teachers indicated that classroom teachers did not comment on the *Art, Books, and Creativity* curriculum, and others indicated that the classroom teachers said the writing component fit well with what they were already doing. Some teachers thought the children's writing improved over the project. Others did not see a change.

Overall, 93 percent of the twenty-seven teachers who responded to the teacher survey agreed or strongly agreed that *Art, Books, and Creativity* supports their district's visual arts learning objectives.

Quality of Professional Development Workshops

Teacher survey responses and focus group results were used to answer the research question "Are the *Art, Books, and Creativity* professional development workshops appropriate and effective?" Overall, 93 percent of the twenty-seven teachers who responded to the teacher survey agreed or strongly agreed that the professional development workshops were useful. Teachers in Albuquerque, who it should be noted were classroom teachers and not art teachers, indicated that the training was very inspiring.

Process and Implementation Changes

Several process and implementation issues were identified in Year One through the use of both formal evaluation tools as well as through less formal means. One issue had to do with scheduling the *Art, Books, and Creativity* post-test at a time that overlapped with state testing, which was found to be a problem because the students were fatigued and disinterested in participating in another test. The second issue had to do with a perceived disconnect between two parts of the curriculum; the introduction of visual arts concepts and vocabulary in the initial lessons and the creation of artists' books in later lessons. The

third issue of concern was that in the first year of the program the experimental and control group score gains on the essay component were not significantly different.

To resolve the test scheduling issue, the post-test was scheduled to take place earlier in the year. Teachers confirmed that the revised post-test schedules did not overlap with their annual state testing.

To address the perceived disconnect within the curriculum, and to address the lack of difference in the score gains of the experimental and control group students, teachers were encouraged to more fully implement the program's writing component. This strategy seemed to be successful in resolving both of those issues.

Summary and Discussion

The National Museum of Women in the Arts received a grant from the U.S. Department of Education in 2003 to develop an arts integration curriculum, *Art, Books, and Creativity*, and to study the impact of that curriculum on student learning. The objectives of *Art, Books, and Creativity* were to increase students' knowledge of visual arts concepts and vocabulary, to increase students written expression related to art concepts, and to increase students' artistic self-confidence. The *Art, Books, and Creativity* program objectives were designed to support the existing learning objectives of participating school districts. Teachers participating in *Art, Books, and Creativity* were provided with professional development workshops and resources designed to support the program objectives. Resource materials included the *Art, Books, and Creativity* curriculum, an *Exploring Art* booklet for each student that served as an introduction to the visual arts and to art museums, a blank journal for each student, art materials for curriculum lessons, resources for each class to visit an art museum, and a series of four visiting artists/writers in each participating classroom.

The learning gains of program participants were tracked and compared to those of non-participants. Learning gains were measured as the difference between pre-test and post-test scores on an art instrument tailored to reflect the *Art, Books, and Creativity* program goals and objectives. These score differences were quantitatively analyzed using a multilevel analysis. Two school districts participated in *Art, Books, and Creativity*: Albuquerque Public Schools in Albuquerque, N.Mex., and Arlington Public Schools in Arlington, Va. In Albuquerque, the program was implemented by classroom teachers and in Arlington the program was implemented by art teachers. In both school districts, all experimental and control schools selected for the program were classified as Title 1 schools so that program effects could be assessed. Over the two-year duration of the program, a total of twenty-five schools, forty-six teachers, and sixty-one classrooms participated in *Art, Books, and Creativity*. An equal number of non-participating schools and teachers acted as a control group. Control schools were matched as closely as possible to the participating schools in terms of student socioeconomic status and general achievement scores.

Over the two-year period, the learning gains in visual arts concepts and vocabulary, written expression, and the composite gains of both components were overall significantly greater for the program participants than those of the non-participating control group. However, in Year One alone, the gains in written expression were not significantly greater for participants, and it was surmised that encouraging fuller use of the program's writing component would remedy this. During the second year of the program, the writing component was more fully implemented than during the first year, and the quantitative analysis indicated that written gains were significantly greater for the participants. The analysis also found that the program effect for this writing component was greater in participating schools with students of lower socioeconomic status.

Teacher focus groups and surveys were used to evaluate the *Art, Books, and Creativity* curriculum, the professional development workshops, the *Exploring Art* text and other resources, the museum visit, the visiting artists' workshops, and the students' creation of artists' books. The professional development workshops were perceived to be both useful and effective by all teachers participating in the program. The training needs of art teachers and general classroom teachers differed, as was expected, and served general classroom teachers particularly well. *Exploring Art* was popular with students, although teachers had mixed opinions about it. Some teachers were concerned that the reading level was too challenging for their particular group of fourth graders; other teachers did not feel the reading level was an issue.

All students participating in *Art, Books, and Creativity* visited an art museum at a mid-point in the program. Students in Arlington visited the National Museum of Women in the Arts in Washington, D.C.; students in Albuquerque visited the Albuquerque Museum of Art in Albuquerque, N.Mex. Teachers reported that students were positively engaged with docents and that they were enthusiastic in demonstrating and applying the visual arts knowledge they had acquired during the first half of the program.

After the museum visit, students met with visiting artists and writers for instruction in creative writing, illustration techniques, and bookmaking. The students then planned and produced their own handmade artists' books. Teachers indicated that students were inspired by the visiting artists and were engaged in the bookmaking activities. The classroom teachers appreciated having a project that integrated language arts and visual arts. They stressed that the project played to the students' strengths, allowing students who struggle to express themselves in either writing or in visual art an opportunity to create something and then respond to it using the other discipline. Not all of the art teachers were comfortable supervising the language arts and writing components of the project, so in the second year many of the art teachers coordinated with their students' classroom teachers on the writing component. In general, teachers perceived that the *Art, Books, and Creativity* program supported their arts education learning objectives as well as language arts learning objectives. In particular, teachers in Albuquerque stressed that the program provided students with more general knowledge and felt the program engaged their students and provided them with motivation to write.

In conclusion, *Art, Books, and Creativity* achieved its main objectives of increasing students' knowledge of visual arts concepts and vocabulary and improving students' written expression related to art concepts. Teachers' feedback pertaining to the *Art, Books, and Creativity* curriculum, professional development workshops, and resources were positive, and teachers indicated that *Art, Books, and Creativity* supports district arts education learning objectives and some language arts learning objectives.

Appendices

Appendix A: School Demographics

Table 6. Albuquerque Year One Experimental and Control Group School Demographics

Experimental Schools	ELL Students	FARMS 2003–2004	Language[1] (median pass rate)	Control Schools	ELL Students	FARMS 2003–2004	Language[1] (median pass rate)
Carlos Rey	73%	84.6%	39	Eugene Field	41%	93.2%	43
East San Jose	85%	97.4%	28	La Mesa	54%	99.1%	34
Eubank	13%	88.1%	34	Lavaland	26%	88.8%	35
Kit Carson	15%	91%	35	Los Padillas	20%	92.1%	37
Longfellow	19%	73.5%	56	Hodgin	26%	79.9%	58
Tomasita	14%	75.6%	50	Montezuma	13%	53.1%	49
Wherry	28%	58.9%	38	Hawthorne	26%	85.9%	40
Experimental Average	**35.9%**	**81.3%**	**40**	**Control Average**	**29.3%**	**84.5%**	**42.3**

[1] Language scores are Terra Nova median pass rates for third grade students.

Table 7. Arlington Year One Experimental and Control Group School Demographics

Experimental Schools	FARMS 2003–2004	English Pass Rate 2002–2003	Control Schools	FARMS 2003–2004	English Pass Rate 2002–2003
Arlington Traditional	12.68%	97%	McKinley	13.49%	100%
Barcroft	62.86%	82.1%	Campbell	76.36%	82.6%
Barrett	72.09%	45.8%	Randolph	76.54%	56.5%
Claremont	41.90%	N/a	Drew	44.26%	74.5%
Long Branch	34.96%	78.7%	Glebe	34.97%	82.4%
Oakridge	62.18%	71.1%	Henry	54.74%	82.8%
Experimental Average	**47.78 %**	**74.9%**	**Control Average**	**50.6%**	**79.8%**

Table 8. Albuquerque Year Two Experimental and Control Group School Demographics

Experimental Schools	ELL[1] Students	FARMS	Language[2]	Control Schools	ELL[1] Students	FARMS	Language[2]
Eugene Field	39.7%	93%	39%	Delores Gonzales	26.3%	99%	39%
La Mesa	46.3%	99%	35%				
Lavaland	33.1%	92%	39%	Emerson	31.9%	97%	34%
Los Padillas	19.9%	93%	37%				
Hodgin	8.2%	71%	65%	Acoma	1.9%	57%	65%
Montezuma	14.1%	47%	51%	S.R. Marmon	8.9%	56%	50%
Hawthorne	24.6%	88%	44%	Eubank	13.0%	88%	84%
Experimental Average[3]	**26.6%**	**83%**	**44%**	**Control Average[3]**	**16.0%**	**79%**	**54%**

[1]ELL student rates are estimates.
[2]Language scores are Terra Nova median percentiles for third grade students. These are the 2003–2004 scores; the 2004–2005 scores were not publicly available at the time the sample was selected.
[3] Average values are based on non-rounded data; data in tables are rounded.

Table 9. Arlington Year Two Experimental and Control Group School Demographics

Experimental Schools	FARMS	English Pass Rate	Control Schools	FARMS	English Pass Rate
Carlin Springs	81%	56%	Randolph	71%	43%
Glebe	34%	66%	Science Focus	30%	89%
Henry	59%	66%	Abingdon	67%	47%
Taylor	6%	88%	Nottingham	5%	87%
Hoffman Boston	76%	59%	Campbell	74%	76%
Experimental Average[1]	**51%**	**67%**	**Control Average[1]**	**49%**	**68%**

[1] Average values are based on non-rounded data; data in tables are rounded.

Appendix B: Essay Rubric

Table 10. Essay Rubric

Vocabulary and Concepts: color (primary, secondary, complementary, intense, bright, low-intensity, neutral, shades, tints, value), shape (organic, geometric), pattern (repeating shapes or colors), space (foreground, background, viewpoint), setting, balance (symmetry), portrait (details, symbols), pose, abstraction (exaggerated color, simplified shapes, few details)					
Reference Back to Work: making the linkage between the vocabulary and concepts and what can be seen in the picture					
Expressive Language: the use of language to convey feelings or attitudes					
Total Points	**Vocabulary/Concepts**		**Reference to Work**		**Expressive Language**
0	**0**		**0**		**0**
1	**1**	**or**	**1**	**or**	**1**
2	**1+**	**and**	**1**	**or**	**1**
3	**2+**	**and**	**1**	**and**	**1**
4	**2+**	**and**	**2+**	**and**	**2+**
A rating of **0** indicates that nothing was written, the writing was off topic, or the writing was indecipherable.					
A rating of **1** indicates that at least one vocabulary word/concept had been discussed, or the student had made reference to the work or had written a reflective interpretation of the work.					
A rating of **2** indicates that at least one vocabulary word/concept had been discussed and had been referenced back to the work or used in a reflective interpretation of the work.					
A rating of **3** indicates that at least two vocabulary words/concepts had been discussed and at least one had been referenced back to the work, and that the student has linked one reflective interpretation to at least one vocabulary word/concept discussed.					
A rating of **4** indicates that two or more vocabulary words/concepts had been discussed, referenced to work, and reflected in expressive language.					

Appendix C: Supplementary Rubric Guidelines

Table 11. Supplementary Rubric Guidelines for Pre-test

Concept	Sub-concept/ Keywords	Examples of References Back to Work	Examples of Interpretation
Color	Color, primary, low intensity, neutral, tints and shades, value	White color on tree branches suggests snow. Red curtains, blue shirt on woman, and yellow cake or bread are primary colors. Yellow and red birds are primary colors.	Low-intensity colors give a sense of quiet and peacefulness or emotion, solitude. The setting seems to be wintertime after a snowstorm; there is snow on the trees but the sky is blue. The table is covered with a white cloth so that the figure (in dark blue and with brown hair) stands out in contrast.
Shape	Organic, geometric	Repeated circular shapes on table. Objects are mainly natural, organic shapes.	
Space	Foreground, background, viewpoint	The woman and her table are inside – in the foreground. The birds and trees are outside, seen through a window – in the background. Our viewpoint is slightly above her, making her seem small in comparison.	The woman's meal is laid out on the table, and she seems to be watching the birds that may also be eating. The items on table are seen from varying perspectives (plates from above, coffee pot from the side).
Setting	Setting	Setting is both inside and outside; interior has table, chair, and dishes. Outside, lack of leaves on trees, suggestion of snow on the branches. Contrast of a single person inside with several birds outside.	Set in wintertime. Separation of the person from the activity outdoors could signify loneliness and isolation, or feelings of calm, enjoyment, warmth of being indoors watching wintry activity outdoors. Interior could be set in a dining room or kitchen.
Balance	Balance, symmetry	Composition is symmetrically balanced. Red curtains on either side. Window frame dividing image in half. Figure dividing table in half. Objects on table in balance.	Balanced composition gives a sense of quiet and calm. Parted curtains are like stage curtains, with a story about to unfold.
Portrait	Details, symbols	Person has her back to us; she has short hair and looks small.	We can't see her face or tell much about her, such as her age or facial expression. Short hair and height may indicate youth.
Pose	Pose	The person has her back toward us, facing the window. She seems small, as if she can only see out of the bottom part of the window.	We can't see her face, so we don't know exactly what she is thinking or who she is. She could be young or old; her height may indicate she is young. She could be watching the birds.
Elements of Abstraction	Abstract, abstraction	The painting is slightly abstract; details are eliminated, outlines are placed around forms, shapes are simplified.	Because the forms are slightly abstract, we have to imagine what the woman is eating.

Table 12. Supplementary Rubric Guidelines for Post-test

Concept	Sub-concept/ Keywords	Examples of References Back to Work	Examples of Interpretation
Color	Color, bright colors, high intensity, complementary, primary, secondary	Exaggerated color. The man's face is partly green. Primary colors (yellow and blue) and secondary colors (green and orange) are used in the painting. Complementary colors (blue and orange, red and green) are also used for emphasis.	Vibrant colors give feeling of energy, could suggest the man's personality is upbeat. The bright colors and varied patterns make the painting lively, although the man is a seated, stable figure. His cool blue jacket stands out in contrast to his warm colored surroundings.
Shape	Organic, geometric	There are both organic (vases of flowers) and geometric shapes (pattern in mesh tabletop, man's tie).	
Pattern	Pattern, repetition, repeating shapes and colors	Table top grid/mesh, pattern in man's tie, shapes in background, patterns of color in background.	Shapes in background provide sense of rhythm, motion. Many different patterns throughout the work make it feel lively, active.
Space	Background, foreground	A variety of objects are on the table in the foreground, such as bowls and vases of flowers. There are shapes in the background. Some plants or other objects are on the right side.	Objects in background and on the table in the foreground could symbolize something about the sitter. There are equally bright (intense) colors in both foreground and background, so space seems shallow.
Setting	Setting	Plants, flowers, table, chair.	Setting suggests he is inside, may be in a room of some sort.
Portrait	Details, symbols	Jacket, tie, white shirt, cigar.	He is well dressed. Clothing suggests man is working, taking a break from work, or perhaps waiting for something/someone.
Pose	Pose, posture	The man is seated, has folded arms. He is holding a cigar.	Sitting posture and facial expressions could indicate he feels relaxed, content. Maybe he is waiting for something. Pose is informal.
Elements of Abstraction	Abstract, abstraction	Exaggerated color on man's face, patterns of colors and shapes in background, large areas of color in background, background seems more abstracted than the man and objects in foreground.	Exaggerated color on man's face could express an emotion, could be used for emphasis or contrast to pink area behind his head. We don't know what the shapes in the background signify.

Appendix D: Mean Achievement Gains Broken Down by Group and Year

Table 13. Mean Achievement Gains Broken Down by Group and Year

Year	Group		Gain: Multiple-choice	Gain: Essay	Gain: Composite
Year One	**Control**	Mean	1.28	0.51	1.99
		N	324	261	261
		Std. Deviation	3.33	1.11	3.50
	Experimental	Mean	2.83	0.77	3.60
		N	432	321	321
		Std. Deviation	3.28	0.80	3.42
	Total	Mean	2.17	0.65	2.88
		N	756	582	582
		Std. Deviation	3.39	0.96	3.55
Year Two	**Control**	Mean	0.54	0.22	0.69
		N	300	276	269
		Std. Deviation	3.41	0.78	3.59
	Experimental	Mean	2.51	0.46	3.10
		N	460	377	372
		Std. Deviation	3.38	0.80	3.49
	Total	Mean	1.73	0.36	2.09
		N	760	653	641
		Std. Deviation	3.52	0.80	3.73
Overall	**Control**	Mean	0.92	0.36	1.33
		N	624	537	530
		Std. Deviation	3.39	0.96	3.60
	Experimental	Mean	2.67	0.60	3.33
		N	892	698	693
		Std. Deviation	1.28	0.51	1.99
	Total	Mean	324	261	261
		N	3.33	1.11	3.50
		Std. Deviation	2.83	0.77	3.60

Appendix E: Statistical Analysis-Multilevel Models for Overall Data

Multiple-choice Component Results

The multiple-choice gain is the difference between the score on the multiple-choice post-test and the score on the multiple-choice pre-test component of the assessment tool. To determine if the multiple-choice score gains for the *Art, Books, and Creativity* participants (experimental group) at the two sites were significantly different from the multiple-choice score gains of the non-participants (control group), the following model was estimated:

Level-1 Model

Multiple-choice score gains = B0 + B1*(GROUP) + R

Level-2 Model

B0 = G00 + U0

B1 = G10

The students in the *Art, Books, and Creativity* experimental group had significantly higher gains on the multiple-choice score than the students in the control group. The mean pre-test and post-test scores for experimental and control groups are presented in Table 14 in raw score form. The mean gains are presented in the body of this report in Table 3 and in Table 13.

Table 14. Overall Average Multiple-choice Pre-test and Post-test Scores

Group		Multiple-choice Pre-test	Multiple-choice Post-test
Control	Mean	9.06	10.03
	N	530	530
	Std. Deviation	3.25	3.69
Experimental	Mean	8.98	11.71
	N	693	693
	Std. Deviation	3.29	3.62
Total	Mean	9.01	10.98
	N	1223	1223
	Std. Deviation	3.28	3.74

The socioeconomic status of the schools, as measured by the percentage of students who receive FARMS, was not a significant predictor, nor was site (Arlington versus Albuquerque). There was significant variation in composite score gains from school to school.

The data were hierarchically structured, with students nested within classes, and classes nested within schools. The sample sizes of individual classes were not sufficient for this

three-level analysis, so a two-level analysis was used instead. Excerpts from the Hierarchical Linear Modeling (HLM) software output relevant to the conclusions presented in this analysis are presented on the next page.

Printout Excerpts: Multiple-choice Gains

The maximum number of Level-1 units = 1264
The maximum number of Level-2 units = 44

Level-1 Model
Y = B0 + B1*(GROUP) + R

Level-2 Model
B0 = G00 + U0
B1 = G10

Sigma_squared = 0.69523

Tau	
INTRCPT, B0	0.8044
Tau (as correlations)	
INTRCPT1,B0	1.000

Random level-1 coefficient	Reliability estimate
INTRCPT1, B0	0.721

The outcome variable is Multiple-choice score gains.

Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard Error	T-ratio	Approx. d.f.	P-value
For INTRCPT1, B0					
INTRCPT2, G00	2.133826	0.193480	11.029	43	0.000
For **GROUP** slope, B1					
INTRCPT2, G10	1.756697	0.384642	4.567	1262	**0.000**

The outcome variable is GAINMC

Final estimation of variance components:

Random Effect	Standard Deviation	Variance Component	df	Chi-square	P-value
INTRCPT1, U0	1.08763	1.18294	43	183.34466	0.000
Level-1, R	3.30193	10.90275			

Statistics for current covariance components model

Deviance	= 6665.377482
Number of estimated parameters	= 2

Essay Score Results

The essay gain is the difference between the score on the essay post-test and the score on the essay pre-test component of the assessment tool. To determine if the essay score gains for the *Art, Books, and Creativity* participants (experimental group) at the two sites were significantly different from the essay score gains of the control groups at the two sites, the following multilevel model was estimated:

Level-1 Model

Essay score gains = B0 + B1*(GROUP) + B2*(YEAR) + R

Level-2 Model

B0 = G00 + U0
B1 = G10 +
B2 = G20 +

The students in the *Art, Books, and Creativity* experimental group had significantly higher gains on the essay score than the students in the control group. The overall average essay pre-test and post-test scores are presented in raw score form in Table 15. The average essay gains are presented in the body of this paper in Table 3 and in Table 13.

Table 15. Overall Average Essay Pre-test and Post-test Scores[1]

Group		Essay Pre-test	Essay Post-test
Control	Mean	1.10	1.47
	N	530	530
	Std. Deviation	0.79	0.69
Experimental	Mean	1.13	1.73
	N	693	693
	Std. Deviation	0.56	0.75
Total	Mean	1.12	1.62
	N	1223	1223
	Std. Deviation	0.67	0.73

[1]The maximum possible essay score is 4.

In the overall analysis, both group and year of implementation were significant predictors of score gain. Students in the program had higher scores than control group students. In addition, students had higher gains in the second year than they did in the first year. This makes sense, because the writing component was more fully implemented during the second year than it was during the first year. Site (Albuquerque versus Arlington) was not a significant predictor of score gains. There was significant variation from school to school.

In the overall analysis, the socioeconomic status of the schools, as measured by the percentage of students enrolled in FARMS, was not a significant predictor of slope, as was the case when looking at Year Two data only. Because the writing component was more fully implemented in Year Two, it is worth reviewing the Year Two finding in this report: in Year Two alone, the program effect was greater for students in schools with higher FARMS rates.

The data were hierarchically structured, with students nested within classes, and classes nested within schools. The sample sizes of individual classes were not sufficient for this three-level analysis, so a two-level analysis was used instead. The relevant printouts from the overall analysis, as well as the Year Two analysis printouts are attached.

Printout Excerpts: Essay Gains Overall

The maximum number of Level-1 units = 1264
The maximum number of Level-2 units = 44

Level-1 Model
Y = B0 + B1*(GROUP) + B2*(YEAR) + R

Level-2 Model
B0 = G00 + U0
B1 = G10
B2 = G20

Sigma_squared = 0.69524
Tau
INTRCPT1, B0 0.06279
Tau (as correlations)
INTRCPT1, B0 1.000

Random Level-1 coefficient	Reliability estimate
INTRCPT1, B0	0.685

Fixed Effect	Coefficient	Standard Error	T-ratio	Approx. d.f.	P-value
For INTRCPT1, B0					
INTRCPT2, G00	0.508953	0.046123	11.035	43	0.000
For **GROUP** slope, B1					
INTRCPT2, G10	0.237069	0.091408	2.594	1261	**0.010**
For **YEAR** slope, B2					
INTRCPT2, G20	-0.276225	0.091765	-3.010	1261	**0.003**

Final estimation of variance components:

Random Effect		Standard Deviation	Variance Component	df	Chi-square	P-value
INTRCPT1,	U0	0.25057	0.06279	43	137.85327	0.000
Level-1,	R	0.83381	0.69524			

Statistics for current covariance components model
Deviance = 3186.387436
Number of estimated parameters = 2

Printout Excerpts: Essay Gains—Year Two

The maximum number of Level-1 units = 814
The maximum number of Level-2 units = 20
Run-time deletion has reduced the number of Level-1 records to 653

Level-1 Model
Y = B0 + B1*(GROUP) + R

Level-2 Model
B0 = G00 + U0
B1 = G10 + G11*(FARMS)

Sigma_squared = 0.57199
Tau
INTRCPT1,B0 0.05109
Tau (as correlations)
INTRCPT1,B0 1.000

Random Level-1 coefficient	Reliability estimate
INTRCPT1, B0	0.708

Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard Error	T-ratio	Approx. d.f.	P-value
For INTRCPT1, B0					
INTRCPT2, G00	0.372386	0.060094	6.197	19	0.000
For **GROUP** slope, B1					
INTRCPT2, G10	0.261771	0.121192	2.160	650	**0.031**
FARMS, G11	0.010507	0.004392	2.392	650	**0.017**

Final estimation of variance components:

Random Effect		Standard Deviation	Variance Component	df	Chi-square	P-value
INTRCPT1,	U0	0.22604	0.05109	19	61.80984	0.000
level-1,	R	0.75630	0.57199			

Statistics for current covariance components model

Deviance	= 1526.513287
Number of estimated parameters	= 2

Composite Results

The composite score gain is the difference between the composite score on the multiple-choice post-test and the score on the multiple-choice pre-test component of the assessment tool. To determine if the composite score gains for the *Art, Books, and Creativity* participants (experimental group) at the two sites were significantly different from the composite score gains of the non-participants (control group), the following model was estimated:

Level-1 Model

Composite Score Gains = B0 + B1*(GROUP) + R

Level-2 Model

B0 = G00 + U0
B1 = G10

The students in the *Art, Books, and Creativity* experimental group had significantly higher gains on the composite score than the students in the control group. The overall average composite pre-test and post-test scores are presented in Table 16 in raw score form. The composite gains are presented in the body of this report in Table 3 and also in Table 13.

Table 16. Overall Average Composite Pre-test and Post-test Scores

Group		Composite Pre-test	Composite Post-test
Control	Mean	10.1623	11.4925
	N	530	530
	Std. Deviation	3.47827	4.04102
Experimental	Mean	10.1111	13.4401
	N	693	693
	Std. Deviation	3.49369	3.96239
Total	Mean	10.1333	12.5961
	N	1223	1223
	Std. Deviation	3.48568	4.11003

The data were hierarchically structured, with students nested within classes, and classes nested within schools. The sample sizes of individual classes were not sufficient for a three-level analysis, so a two-level analysis was used instead. Excerpts from the Hierarchical Linear Modeling (HLM) software output relevant to the conclusions presented in this analysis are presented on the next page.

Printout Excerpts: Overall Composite Gains

The maximum number of Level-1 units = 1264
The maximum number of Level-2 units = 44

Level-1 Model
Y = B0 + B1*(GROUP) + R

Level-2 Model
B0 = G00 + U0
B1 = G10

Sigma_squared = 12.29006
Tau
INTRCPT1,B0 1.51595
Tau (as correlations)
INTRCPT1,B0 1.000

Random level-1 coefficient	Reliability estimate
INTRCPT1, B0	0.745

The outcome variable is Composite Score Gains

Final estimation of fixed effects:

Fixed Effect	Coefficient	Standard Error	T-ratio	Approx. d.f.	P-value
For INTRCPT1, B0					
INTRCPT2, G00	2.660886	0.215507	12.347	43	0.000
For **GROUP** slope, B1					
INTRCPT2, G10	1.962840	0.427795	4.588	1262	**0.000**

Final estimation of variance components:

Random Effect		Standard Deviation	Variance Component	df	Chi-square	P-value
INTRCPT1,	U0	1.23124	1.51595	43	193.35018	0.000
level-1,	R	3.50572	12.29006			

Statistics for current covariance components model

Deviance	= 6820.482704
Number of estimated parameters	= 2