

Museum Visitor Studies, Evaluation and Audience Research

Randi Korn & Associates, Inc.
118 East Del Ray Avenue
Alexandria, VA 22301

**Educational Research:
Evaluation of *Traveling Through Time*,
a School Program of the Museum of the
City of New York**

VOLUME 1: REPORT

Prepared for the
Museum of the City of New York
New York, NY

TABLE OF CONTENTS

LIST OF TABLES	IV
LIST OF FIGURES	V
ABSTRACT	VI
EXECUTIVE SUMMARY	VII
Introduction.....	vii
Student Characteristics	vii
Principal Findings: Student Questionnaires	vii
Principal Findings: Student Interviews	viii
DISCUSSION.....	X
Introduction.....	x
Attitudes and Ideas about History and Museums	x
Historical Thinking Skills: A Comparison of Control and Treatment Students.....	xi
Historical Thinking Skills: Overall Achievement	xii
Implications for MCNY.....	xiii
Suggestions for Future Research	xv
References	xv
INTRODUCTION	I
Project History.....	1
Research Design.....	1
Methodology and Analysis.....	3
Reporting Method.....	6
Reliability and Validity of the Study	6
PRINCIPAL FINDINGS: STUDENT QUESTIONNAIRES	7
Introduction.....	7
Student Characteristics	7
History Attitudes and Knowledge	9
Opinions of MCNY Fieldtrip	13

PRINCIPAL FINDINGS: STUDENT INTERVIEWS 15

Introduction 15
Historical Knowledge 19
Historical Inquiry 22
Historical Perspective 28
Historical Reasoning 31
Summary of Treatment Students' Scores 39

APPENDICES VOLUME 2

LIST OF TABLES

TABLE 1:	Methodological Timeline.....	3
TABLE 2:	Students and Schools by Year.....	7
TABLE 3:	Age by Group.....	8
TABLE 4:	Gender by Group.....	8
TABLE 5:	Language(s) Spoken at Home by Group.....	8
TABLE 6:	“Other” Languages Spoken at Home by Group.....	9
TABLE 7:	MCNY Visits by Group.....	9
TABLE 8:	Attitudes about History by Group.....	10
TABLE 9:	Thoughts about History Knowledge by Group.....	11
TABLE 10:	Attitudes about History Museums by Group.....	12
TABLE 11:	Thoughts about Artifacts by Group.....	13
TABLE 12:	Favorite Parts of MCNY Fieldtrip by Group.....	13
TABLE 13:	Opinions of MCNY Fieldtrip.....	14
TABLE 14:	Students by Group.....	15
TABLE 15:	Criteria for Rubric 1a—Historical Knowledge of the Lenape.....	19
TABLE 16:	Criteria for Rubric 1b—Historical Knowledge of the Dutch.....	20
TABLE 17:	Criteria for Rubric 1c—Historical Knowledge of Trade between the Lenape and Dutch.....	21
TABLE 18:	Criteria for Rubric 2a—Historical Inquiry about the Lenape.....	22
TABLE 19:	Criteria for Rubric 2b—Historical Inquiry about the Dutch.....	23
TABLE 20:	Criteria for Rubric 2c—Historical Inquiry about Trade between the Lenape and Dutch.....	24
TABLE 21:	Criteria for Rubric 2d—Historical Inquiry for Observation of the Candle Maker.....	25
TABLE 22:	Criteria for Rubric 2e—Historical Inquiry for Formulating Hypotheses about the Candle Maker.....	26
TABLE 23:	Criteria for Rubric 2f—Historical Inquiry for Object Identification.....	27
TABLE 24:	Criteria for Rubric 3a—Historical Perspective of Differences between the Lenape and Dutch.....	28
TABLE 25:	Criteria for Rubric 3b—Historical Perspective of Different Belief Systems.....	29
TABLE 26:	Criteria for Rubric 3c—Historical Perspective of Cultural Variation.....	30
TABLE 27:	Criteria for Rubric 4a—Historical Reasoning about Why the Lenape and Dutch Traded.....	31
TABLE 28:	Criteria for Rubric 4b—Historical Reasoning about Changing Landscape of New York City.....	32
TABLE 29:	Criteria for Rubric 4c—Historical Reasoning about the Significance of the Lenape to New York City History.....	33
TABLE 30:	Criteria for Rubric 4d—Historical Reasoning about the Significance of the Dutch to New York City History.....	34
TABLE 31:	Criteria for Rubric 4e—Historical Reasoning about the Significance of New Amsterdam to New York City History.....	35
TABLE 32:	Criteria for Rubric 4f—Historical Reasoning about the Significance of Exploration to New York City History.....	36
TABLE 33:	Criteria for Rubric 4g—Historical Reasoning about the Significance of Beavers to New York City History.....	37
TABLE 34:	Criteria for Rubric 4h—Historical Reasoning about the Significance of Trade to New York City History.....	38

LIST OF FIGURES

FIGURE 1:	Picture A.....	16
FIGURE 2:	Picture B.....	16
FIGURE 3:	Picture C.....	17
FIGURE 4:	Candle Maker.....	18
FIGURE 5:	Achievement on Rubric 1a—Historical Knowledge of the Lenape.....	19
FIGURE 6:	Achievement on Rubric 1b—Historical Knowledge of the Dutch.....	20
FIGURE 7:	Achievement on Rubric 1c—Historical Knowledge of Trade between the Lenape and Dutch.....	21
FIGURE 8:	Achievement on Rubric 2a—Historical Inquiry about the Lenape.....	22
FIGURE 9:	Achievement on Rubric 2b—Historical Inquiry about the Dutch.....	23
FIGURE 10:	Achievement on Rubric 2c—Historical Inquiry about Trade between the Lenape and Dutch.....	24
FIGURE 11:	Achievement on Rubric 2d—Historical Inquiry for Observation of the Candle Maker.....	25
FIGURE 12:	Achievement on Rubric 2e—Historical Inquiry for Formulating Hypotheses about the Candle Maker.....	26
FIGURE 13:	Achievement on Rubric 2f—For Historical Inquiry about Object Identification.....	27
FIGURE 14:	Achievement on Rubric 3a—Historical Perspective of Differences between the Lenape and Dutch.....	28
FIGURE 15:	Achievement on Rubric 3b—Historical Perspective of Different Belief Systems.....	29
FIGURE 16:	Achievement on Rubric 3c—Historical Perspective of Cultural Variation.....	30
FIGURE 17:	Achievement on Rubric 4a—Historical Reasoning about Why the Lenape and Dutch Traded.....	31
FIGURE 18:	Achievement of Rubric 4b—Historical Reasoning about Changing Landscapes of New York City.....	32
FIGURE 19:	Achievement on Rubric 4c—Historical Reasoning about the Significance of the Lenape to New York City History.....	33
FIGURE 20:	Achievement on Rubric 4d—Historical Reasoning about the Significance of the Dutch to New York City History.....	34
FIGURE 21:	Achievement on Rubric 4e—Historical Reasoning about the Significance of New Amsterdam to New York City History.....	35
FIGURE 22:	Achievement on Rubric 4f—Historical Reasoning about the Significance of Exploration to New York City History.....	36
FIGURE 23:	Achievement on Rubric 4g—Historical Reasoning about the Significance of Beavers to New York City History.....	37
FIGURE 24:	Achievement on Rubric 4h—Historical Reasoning about the Significance of Trade to New York City History.....	38
FIGURE 25:	Summary of Treatment Students’ Scores.....	39

ABSTRACT

With funding from Atlantic Philanthropies, the Museum of the City of New York (MCNY) contracted Randi Korn & Associates, Inc. (RK&A) to evaluate the *Traveling Through Time (TTT)* program, a 90-minute fieldtrip program facilitated by museum educators that explores the early history of New York City through objects and inquiry. The study explores whether and to what extent the *TTT* program affects students' attitudes about history and history museums, history knowledge, and history-related skills. In the last 30 years, there has been little rigorous research that investigates the outcomes (particularly learning outcomes) of one-time fieldtrip experiences, so this study presented a unique opportunity to fill the knowledge gap.

RK&A administered questionnaires and interviews to a treatment group and control group of fourth-grade students in New York City public schools; the data are quantitative. The study revealed that the single visit fieldtrip, the *TTT* program, had positive effects on students' attitudes and ideas about history as well as positive effects on students' achievement of history-related skills, including historical knowledge, historical inquiry, historical perspective, and historical reasoning. However, the study also revealed that students' achievement of history-related skills overall was modest, indicating that students have not reached their full potential in the classroom and at the Museum.

EXECUTIVE SUMMARY

INTRODUCTION

This report presents the findings from the evaluation of *Traveling Through Time (TTT)*, a school program of the Museum of the City of New York (MCNY). With funding from Atlantic Philanthropies, the Museum contracted Randi Korn & Associates, Inc. (RK&A) to conduct the evaluation, which explored the extent to which the *TTT* program enhanced fourth-grade students' attitudes about history and history museums as well as historical knowledge, inquiry, perspective, and reasoning. The summary below highlights key findings from questionnaires and interviews administered to a treatment group (fourth-grade students who attended *TTT*) and a control group (fourth-grade students who did not attend *TTT*); all data were collected in February and March 2010 at participating New York City public schools.

**Selected highlights of the study are included in this summary.
Please consult the body of the report for a detailed account of the findings.**

STUDENT CHARACTERISTICS

Participating students were from three schools in Spanish Harlem; schools were selected because students were similar in demographics, socioeconomic characteristics, and test scores. Of the sample, more than two-thirds of students were 9 years old (69 percent), and one-half of students were boys (50 percent). Most students said they speak English at home (95 percent), and more than two-thirds also speak Spanish at home (70 percent).

PRINCIPAL FINDINGS: STUDENT QUESTIONNAIRES

METHODOLOGY

A total of 154 fourth-grade students completed questionnaires; about one-half were control students and one-half were treatment students. To administer the questionnaires, a trained data collector read each question aloud to a classroom and students responded on a printed questionnaire.

HISTORY ATTITUDES

- ♦ All students have mostly positive attitudes about history. There is one statistically significant difference between control and treatment students' responses; treatment students responded more favorably than control students to the statement, "Learning history is only important for school" (i.e., strongly disagreed with the statement).
- ♦ All students have mostly positive attitudes about museums about history. There are two statistically significant differences between control and treatment students' responses; treatment students responded more favorably than control students to the statement, "In museums about history, I am not allowed to touch anything" (i.e., strongly disagreed with statement), while control students responded more favorably than treatment students to the

statement, “The best thing about museums about history is the interesting objects” (i.e., strongly agreed with the statement).

- ◆ Treatment students’ thoughts about history knowledge are moderately sophisticated; at least one-quarter indicated that they know what they know about history from primary sources (e.g., objects and maps from the past, old photographs, and talking with parents/grandparents about their life). There are two statistically significant differences between control and treatment students’ responses; treatment students are more likely than control students to indicate that they know what they know about history from “old photographs,” while control students are more likely than treatment students to indicate “teachers.”
- ◆ Treatment students’ thoughts about artifacts are moderately developed, with at least one-half connecting artifacts to the past. There is one statistically significant difference between control and treatment students’ responses; control students are more likely than treatment students to indicate that artifacts are something “you cannot touch.”

OPINIONS OF THE MUSEUM OF THE CITY OF NEW YORK FIELDTRIP

- ◆ Treatment students responded favorably to the statements about the Museum of the City of New York.
- ◆ Treatment students said the best part of their visit to the Museum of the City of New York was talking about the past and learning about New York City.

PRINCIPAL FINDINGS: STUDENT INTERVIEWS

METHODOLOGY

A total of 150 fourth-grade students participated in interviews; about one-half were control students and one-half were treatment students. During the interview, data collectors presented students with images, note cards with key terms, and an object related to early New York City history (e.g., time period of Lenape and Dutch settlement). Interviews were audio recorded to facilitate analysis. RK&A created rubrics to measure students’ achievement in four skill areas: historical knowledge, historical inquiry, historical perspective, and historical reasoning.

HISTORICAL KNOWLEDGE

Historical knowledge is students’ factual knowledge about people or events. The three knowledge areas explored in this study are the Lenape, Dutch, and trade between the Lenape and Dutch.

- ◆ Treatment students scored higher than control students on all three historical knowledge areas measured; the differences are statistically significant.
- ◆ At least three-quarters of treatment students scored at the top half of each knowledge rubric (“Developing” or “Accomplished” level).
- ◆ From among the four skill areas, students scored best on historical knowledge.

HISTORICAL INQUIRY

Historical inquiry explores students’ ability to explain how they know what they know about an object, illustration, or primary source material. The six inquiry areas explored in this study are the Lenape, Dutch, trade between the Lenape and Dutch, observation of a candle maker, formulating hypotheses for the candle maker, and object identification.

- ◆ Treatment students scored higher than control students on five of the six historical inquiry areas measured; the differences are statistically significant.
- ◆ For the most part, one-half of treatment students scored at the top half of each inquiry rubric (“Developing” or “Accomplished” level).
- ◆ Students scored moderately well on historical inquiry (compared to historical knowledge, historical perspective, and historical reasoning).

HISTORICAL PERSPECTIVE

Historical perspective explores students’ understanding of and appreciation for differences among peoples/situations/cultures. The three perspective areas explored in this study are differences between the Lenape and Dutch, difference in belief systems, and cultural variation.

- ◆ Treatment students scored higher than control students on two of the three historical perspective areas measured; the differences are statistically significant.
- ◆ At least one-third of treatment students scored at the top half of each perspective rubric (“Developing” or “Accomplished” level).
- ◆ Students scored moderately well on historical perspective (compared to historical knowledge, historical inquiry, and historical reasoning).

HISTORICAL REASONING

Historical reasoning explores students’ understanding of cause and effect and/or change over time. The six areas of reasoning assessed in the study were: explaining why the Lenape and Dutch traded with one another, explaining changes in landscape over time, and the significance of the Lenape, Dutch, New Amsterdam, trade, beavers, and exploration to the history of New York City.

- ◆ Treatment students scored higher than control students on seven of the eight historical reasoning areas measured; the differences are statistically significant.
- ◆ For the most part, more than one-half of treatment students scored at the bottom half of each rubric (“Below Beginning” and “Beginning”).
- ◆ Students scored lowest on historical reasoning (compared to historical knowledge, historical inquiry, and historical perspective).

DISCUSSION

INTRODUCTION

In 2007, Randi Korn & Associates, Inc. (RK&A) conducted a formative evaluation of *Traveling Through Time (TTT)* and other Museum of the City of New York (MCNY) school programs to identify strengths and weaknesses of the programs and make suggestions for program improvement (RK&A, 2007a). Findings of that evaluation revealed that the programs successfully connected with classroom curriculum and New York State History Standards. Furthermore, the programs highly engaged students as a result of the clear orientation, object-based focus, and skilled staff museum educators who facilitated the programs using inquiry. These positive findings inspired MCNY to pursue further funding to evaluate whether the programs were achieving intended student outcomes.

Thus, with funding from Atlantic Philanthropies, MCNY contracted RK&A in 2009 to conduct a quasi-experimental study measuring the outcomes of one specific program, *TTT*, for fourth-grade students. Findings reveal that the program is successful; in particular, the program positively affected students' attitudes and ideas about history as well as students' achievement of history-related skills. Nevertheless, the study also reveals that students' achievement of history-related skills was modest, which raises questions not only about the program but about the abilities of fourth-grade students in regard to historical thinking. The following discussion explores these two findings in the context of the program and history education. The discussion also explores the implications of these findings and offers suggestions for program improvement and further research.

ATTITUDES AND IDEAS ABOUT HISTORY AND MUSEUMS

Notably, RK&A found that *TTT* does indeed have quantifiable effects on students' attitudes *and* ideas about history. First, findings show that *all* students held positive attitudes about "learning history." This is not surprising, as elementary-age students tend to hold positive attitudes about school and learning; negative attitudes begin to develop around middle school (Anderman & Midgley, 1998; Eccles & Midgley, 1989; Hogsten & Peregoy, 1999). Nevertheless, in one attitude measure a statistically significant difference emerged between treatment and control students.¹ Treatment students are more likely than control students to strongly *disagree* with the statement, "Learning history is only important for school," suggesting that students who participated in *TTT* see the relevance of learning history outside the classroom.

In regard to attitudes toward "history museums," a couple of differences emerged. First, treatment students are more likely than control students to say that one can touch things in history museums. This is not surprising since, in *TTT*, students have an opportunity to handle and talk about several different objects from New Amsterdam (i.e., a bedwarmer and wooden clogs). However, and curiously, control students are more likely than treatment students to say that the best thing about history museums is the objects. One might have expected the opposite to be true since, as just stated, treatment students handled objects in the program. One explanation for this seemingly contradictory finding is that control students *imagine* that touching objects would be the best part of a trip to a history museum, while

¹ Throughout this discussion, "Treatment students" refers to those students who participated in the *Traveling Through Time* program, and "control students" refers to students who did not participate in *Traveling Through Time*.

treatment students, having participated in *TTT*, can make a judgment about what they liked best based on their *actual* experience with *all* the various aspects of the program, which includes discussions, objects, dioramas, and other primary source materials. To better understand this finding, another question from the survey provides more insight. The question (asked only of treatment students since it was specific to the program) listed all the various aspects of *TTT* and asked students to select the two they liked best. Somewhat surprisingly, the students indicated that “talking about the past” and “learning about New York City” was the best part of their Museum visit as compared to “touching objects used in the past,” “making a Delft tile,” “spending time away from school,” and “visiting a new place.” This finding gives us pause since conventional wisdom says that students are most excited by hands-on experiences. While one cannot dispute the power of hands-on experiences, this finding suggests that student discussions facilitated by educators using inquiry can be just as exciting as touching objects or making something, especially when inquiry is executed effectively.

Finally, in regard to students’ ideas about history, or “how we know what we know” about history, treatment students demonstrated a more sophisticated understanding. Case in point, while more control students indicated we know what we know from teachers (e.g., a liaison), more treatment students indicated we know what we know from old photographs (e.g., primary source material). Interestingly, emphasizing to students that we learn history through primary sources is not an explicit goal of the *TTT* program. Nevertheless, we hypothesize that the program’s utilization of a range of sources, including objects, recreated period rooms, maps, and illustrations, conveys that history comes from more places than textbooks and teachers. Further, the program invited treatment students to handle artifacts in the program, which seemingly resulted in treatment students’ greater understanding of artifacts as objects that may be touched and interpreted (as opposed to control students who perceive artifacts as off-limits—something they cannot touch); this, too, likely contributed to treatment students’ more sophisticated understanding about learning history.

HISTORICAL THINKING SKILLS: A COMPARISON OF TREATMENT AND CONTROL STUDENTS

While this study examined attitudes as described above, the primary focus was to determine whether *TTT* had achieved its student outcomes related to historical thinking skills. The first step in the study was to clearly define the student outcomes. A meeting with education staff and a review of literature led to the identification of four distinct skills, each associated with multiple content areas specific to the *TTT* program. Measuring the kind of learning that results from a museum-based program (often referred to as a fieldtrip) is extremely challenging. The effects of a one-time 90-minute program can be subtle and difficult to detect through conventional evaluation strategies. To address this challenge, RK&A developed rubrics to measure attainment of each of the skill areas. A rubric is a set of criteria, linked to learning objectives that is used to assess a performance of knowledge, skills, etc. along a continuum, and in this study, on a scale from 1, “Below Beginning,” to 4, “Accomplished” (see p. 5 for a description of rubric development). The continuum reveals subtle differences in the attainment of skills or understandings.

Overall, findings demonstrate that *TTT* positively affected treatment students' achievement in all of the four skill areas. These four skill areas are described below (for the complete rubric, see Appendix F). Significant differences between treatment and control students are stated within each category.

HISTORICAL KNOWLEDGE is students' factual knowledge about people or events. The three knowledge content areas explored in this study are the Lenape, Dutch, and trade between the Lenape and Dutch. **Treatment students scored higher in all three knowledge content areas.**

HISTORICAL INQUIRY explores students' ability to explain how they know what they know about an object, illustration, or primary source material. The six inquiry content areas explored in this study are the Lenape, Dutch, trade between the Lenape and Dutch, observation of a candle maker, formulating hypotheses about the candle maker, and object identification. **Treatment students scored higher on all inquiry content areas except "formulating hypotheses about the candle maker."**

HISTORICAL PERSPECTIVE explores students' understanding of and appreciation for differences among peoples/situations/cultures. The three perspective content areas explored in this study are differences between the Lenape and Dutch, difference in belief systems, and cultural variation. **Treatment students scored higher on all perspective content areas except "cultural variation."**

HISTORICAL REASONING explores students' understanding of cause and effect and/or change over time. The eight content areas of reasoning assessed in the study were: explaining why the Lenape and Dutch traded with one another, explaining changes in landscape over time, and the significance of the Lenape, Dutch, New Amsterdam, trade, beavers, and exploration to the history of New York City. **Treatment students scored higher on all reasoning content areas except "explaining changes in landscape over time."**

HISTORICAL THINKING SKILLS: OVERALL ACHIEVEMENT

Not only does the rubric allow us to explore whether treatment students' historical thinking skills are more developed than those of control students, it also allows us to quantify and explore to what extent students achieved the skills against what the Museum considers ideal. Findings show that despite great differences between the treatment students and control students, overall achievement of the history skills measured was modest, with treatment students often scoring at the middle of the rubric. This finding indicates that the students have room for improvement.

Looking closely at the findings, we recognize that students are performing well in historical knowledge (three-quarters of treatment students scored at the top half of the rubric).² *TTT* helped students gain greater knowledge of who the Lenape and Dutch were as well as the nature of the trade that occurred between them. This finding is not particularly surprising since factual identification is a lower level

² When we say that students scored at the top half of the rubric, we mean that they scored at either the "Developing" or "Accomplished" level.

cognitive skill than the other three skills measured. Regardless, this gain is notable as it aligns with the New York State History Standards.

However, students did not perform as well in the other three skill areas. In historical inquiry, just one-half of treatment students scored at the top half of the rubric. It may be surprising that treatment students did not score higher in this area, especially since findings from the questionnaire demonstrate that treatment students have a greater understanding that historical knowledge is embedded in primary source materials. We can speculate an explanation for students' moderate scores in this area. For instance, being able to interpret an illustration or object is something that must be cultivated over time by parents and teachers. If students only exposure to inquiry- and object-based learning was in the *TTT* program, it is likely that interpreting a primary source or illustration was a novel experience, explaining why gains in this area were modest.

In historical perspective and historical reasoning, the majority of students scored at the bottom half of the rubric. It is not surprising that students scored relatively low in these two areas since they both are high-level cognitive skills, and research is conflicted as to whether elementary students have the developmental capacity to attain these skills (Brophy & VanSledright, 1997); nevertheless, it is disappointing and somewhat alarming to the Museum and educators because it suggests that students are unable to “apply” knowledge. For instance, while students were able to identify the Lenape and Dutch, they were less able to describe how the Lenape and Dutch are different from one another or explain their significance to the history of New York City. This finding alerts us to a potential flaw in the way that history is traditionally taught. Yet, the findings from this study strongly suggest that with repeated exposure to programs like *TTT*, students can develop skills like historical reasoning and perspective. Proof positive, the fact that treatment students outscored control students in these areas suggests that aspects of *TTT* help students apply and think critically about their knowledge, and we propose that these aspects include the dialogue that happens during the program, the use of objects, and the inquiry strategies employed by the museum educators. While teachers and museum professionals have been working to reform the way history is taught over the last couple decades, there are still significant barriers to integrating dialogue, inquiry, and interpretation into the curriculum (Barton & Levstik, 2003).

IMPLICATIONS FOR MCNY

The fact that *TTT* affected students so positively across so many areas is remarkable, especially when one considers that treatment students were only exposed to a 90-minute program. Unfortunately, we cannot compare these findings to other studies because fieldtrips and school programs across museums are highly unique, and it is difficult to verify whether one program experience is relatively equivalent to another. However, we know that changes in knowledge and understanding typically require repeated exposure. It is impossible to know exactly what about the program is so effective. Yet, one can speculate, especially based on the 2007 formative evaluation which found that:

Observation after observation demonstrated that the programs were dynamic, student-centered, interactive, object-based, and exciting. Educators led students from activity to activity, shifting gears from one idea to the next, all while maintaining a constant dialogue with each other. Through open-ended questioning and the use of objects and artifacts designed to make students think critically, museum educators helped students construct knowledge for themselves (RK&A, 2007a).

In summary, the following program characteristics are likely attributable to the achievement of student outcomes:

- ♦ The program is object-based. Museums are the greatest advocates for object-based education, emphasizing the belief that encounters with real objects present the opportunity for education. For instance, objects are great discussion pieces because they offer visitors many entry points (Shuh, 1999). Ideas formulated through an experience with a “real thing” are absorbed, retained, and generate curiosity to know more (Xanthoudaki, 1998). And, Paris (2002) argues that investigations of objects allow children to discover both the natural and cultural history of objects.
- ♦ The program is inquiry-based. As was found in the formative evaluation, *TTT* museum educators skillfully ask open-ended questions to lead and guide the delivery of content (RK&A, 2007a). Museum educators avoid lectures, dichotomous questions, and rarely asked close-ended questions. Modern educational and neuroscience research support this finding; research indicates that dialogue—talking, sharing, and discussing—is a critical component to meaning-making (Jensen, 1998). Further, inquiry itself has proven to be an effective teaching strategy across a range of fields, including art, science, and history (Bruner, 1961; Housen, 2002; National Science Foundation, 1999; RK&A, 2007b). As noted in Jensen (1998), pushing students to answer “how” questions can help “expose the boundaries, limitations, and genius in student thinking” (p. 97).
- ♦ The facilitators of the program are paid museum educators, rather than volunteers. Because they are on the staff of MCNY rather than freelancers or volunteers (which is often the case in museum programs and tours), they are invested and committed to the programs. Not only do the educators lead the programs, but they help to create and shape the nature of the programs.
- ♦ The program has strong ties to curriculum and state standards, a widely accepted element of school program and fieldtrip success (DeWitt & Storksdieck, 2008; Institute for Learning Innovation, 2006).

Despite these positive implications, it is worth looking closely at the findings to identify weaknesses and potential shortcomings of the program, some of which the MCNY may be able to address. These shortcomings are described below:

- ♦ Findings suggest that the program could improve the small group, object-based activity of *TTT*. During *TTT*, students sit in small groups, and museum educators facilitate students in the act of describing, discussing, and speculating about several objects from New Amsterdam. Thus, one would assume treatment students would demonstrate a greater ability to interpret an object; however, findings show that treatment students were no more able to formulate a hypothesis about the candle maker than were control students. This finding is surprising given the content of the program. It may be that the program is effective in helping students with their observation skills (findings do indeed show that treatment students were better able to *describe* the candle maker than control students); however, the program falls short in helping students with their interpretation skills.
- ♦ Findings suggest that the program could better help students understand historical perspective, that is, appreciate the differences between cultural groups (specifically, the Dutch and Lenape). For instance, treatment students and control students scored nearly identical in the area of “cultural variation,” which describes students’ acceptance of differences between the Lenape and Dutch. Nearly all the students scored at the developing level, meaning that in describing the different people in an image (the Dutch and the Lenape), most students stated differences as facts rather than judgments. This finding is no

doubt positive; however, MCNY intends for students to see beyond those differences and *explicitly* understand that neither the Dutch nor the Lenape are better than one another. Furthermore, all students scored very low in their ability to describe the different belief systems of the Lenape and Dutch; in fact, most students demonstrated no knowledge of the differences in ideas about land ownership or resources.

- ◆ Findings suggest that the program could better help students develop historical reasoning skills. Except for being able to explain why the Dutch and Lenape traded, most treatment and control students scored relatively low in the area of historical reasoning. Of critical importance, however, is that treatment students scored significantly higher than control students in nearly every area of historical reasoning. As indicated earlier, this suggests that the program does indeed have an effect on students' reasoning skills, but could do better.
- ◆ Creating a Delft tile does not seem to resonate with the students; it ranked lowest when students were asked what they liked best about the program. While the activity is designed to be a reflective activity, which has the potential to be a highly effective way to further learning, the activity may be better facilitated as a post-visit activity for the classroom versus at the Museum, where students are having highly engaging conversations about history in a novel setting.

SUGGESTIONS FOR FURTHER RESEARCH

While research on fieldtrips in general is a growing body of research, one challenge is that fieldtrips are conducted in a variety of ways and in a variety of settings, so it is difficult to look at the literature as a whole without sufficient context for the programs and the methods and rigor of the studies. This study demonstrates the effectiveness of an object-based museum fieldtrip on students and must be considered in this context. Nevertheless, the study poses interesting questions and considerations for future research:

- ◆ What correlations are there between students' understanding of how we know what we know about history, their attitudes about history, and their achievement of history-related skills?
- ◆ Why did students have difficulty applying their knowledge to reasoning? How can museum educators more effectively help students build their ability to reason?
- ◆ This study measured students' knowledge gain shortly after the fieldtrip, allowing researchers to consider the results in the context of museum practice, which is a very useful evaluation strategy. What would have happened if the study was conducted after a few weeks, months, or even years? While exploring the longer-term effects of fieldtrips might be beneficial, such research is prohibitive in terms of cost and logistics (DeWitt & Storksdieck, 2008).

REFERENCES

- Anderman, L. H., & Midgley, C. (1998). *Motivation and middle school students*. Champaign, IL: ERIC Clearinghouse on Elementary and Early Childhood Education.
- Barton, K. C., & Levstik, L. S. (2003). Why don't more history teachers engage students in interpretation? *Social Education*, 67(6), 358-361.
- Bruner, J. S. (1961). The act of discovery. *Harvard Educational Review*, 31(1), 21–32.

- Brophy, J. E., & VanSledright, B. (1997). *Teaching and learning history in elementary schools*. New York, NY: Teachers College, Columbia University.
- Eccles, J. S., & Midgely, C. (1989). Stage-environment fit: Developmentally appropriate classrooms for early adolescents. In R. E. Ames & C. Ames (Eds.), *Research on motivation in education*. New York, NY: Academic Press.
- DeWitt, J., & Storksdieck., M. (2008). A short review of school field trips: Key findings from the past and implications for the future. *Visitor Studies*, 11(2), 181-197.
- Hogsten, J. F., & Peregoy, P. A. (1999). *An investigation of reading attitudes and self-perceptions of students reading on or below grade level: Research report*. Charlottesville, VA: University of Virginia.
- Housen, A. (2002). Aesthetic thought, critical thinking, and transfer. *Arts and Learning Research Journal*, 18(1), 99-131.
- Institute for Learning Innovation. (2006). *Results from the quality fieldtrip study: Assessing the LEAD program in Cleveland, Ohio*. Edgewater, MD: Institute for Learning Innovation.
- Jensen, E. (1998). *Teaching with the brain in mind*. Alexandria, VA: Association for Supervision and Curriculum Development.
- National Science Foundation. (1999). *Foundations volume 2: Inquiry—thoughts, views, and strategies for the K-5 classroom*. Washington, D.C.: National Science Foundation.
- Paris, S. G., & Mercer, M. G. (2002). Finding self in objects: Identity exploration in museums. In G. Leinhardt, K. Crowley, & K. Knutson (Eds.), *Learning Conversations in Museums* (401-423). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Randi Korn & Associate, Inc. (2007a). *Program evaluation: School programs of the Frederick A.O. Schwarz Children's Center at the Museum of the City of New York*. Alexandria, VA: Randi Korn & Associates, Inc.
- Randi Korn & Associate, Inc. (2007b). *Synthesis of 2004-05 and 2005-06 studies: Teaching literacy through art (Solomon R. Guggenheim Museum)*. Alexandria, VA: Randi Korn & Associates, Inc.
- Shuh, J. H. (1999). Teaching yourself to teach with objects. In E. Hooper-Greenhill (Ed.), *The Educational Role of the Museum* (80-91). London: Routledge.
- Xanthoudaki, M. (1998). Is it really worth the trip?: The contribution of museum and gallery educational programmes to classroom art education. *Cambridge Journal of Education*, 28(2), 181-195.

INTRODUCTION

The Museum of the City of New York (MCNY) contracted Randi Korn & Associates, Inc. (RK&A) to study its school program *Traveling Through Time (TTT)*. This report presents findings from a quasi-experimental research study that explored the extent to which the program enhanced fourth-grade students' attitudes about history and museums about history as in four skill areas—historical knowledge, historical inquiry, historical perspective, and historical reasoning as related to the history of New York City. The study was made possible by Atlantic Philanthropies.

PROJECT HISTORY

Traveling Through Time is a 90-minute program that takes place at MCNY. The program—available to K-8 students—is facilitated by a MCNY educator. In the program, students learn how and why New Amsterdam became New York as they tour the *Trade* and *New York Interiors* exhibitions. Children also create their own “Delft tiles” to take home. Throughout the program, MCNY educators employ inquiry and object-based education strategies.

Over its long history, the *TTT* program has undergone many transformations. While MCNY has evaluated the program over its various iterations and the previous assessments revealed some of the successes and challenges of the program, this study is the first to explore the effects that *TTT* has on students.

RESEARCH DESIGN

OBJECTIVES OF THE STUDY

The research plan was designed to measure students' history learning including attitudes, knowledge, and skills. Specifically, the study measured:

- ◆ Whether, and to what extent, *TTT* affects or changes students' attitudes toward history and museums about history;
- ◆ Whether, and to what extent, *TTT* is meeting student outcomes related to historical knowledge, understanding of cause and effect (reasoning), and cultural difference (perspective); and,
- ◆ Whether, and to what extent, *TTT* is meeting student outcomes related to historical inquiry skills.

LITERATURE REVIEW

Prior to and during the research process, RK&A compiled research to contextualize and inform this research study (see Appendix A).

OVERVIEW OF THE STUDY

RK&A employed a quasi-experimental design, designating a treatment group—students who received *TTT*—and control group—students who did not receive *TTT*. Measures included student questionnaires and student interviews. All measures are discussed in detail in the “Methodology and Analysis” section on page 3.

SAMPLING

RK&A sampled 154 students, of which one-half were part of the control group and one-half were part of the treatment group. To eliminate any variables that may affect the data, the study included only New York City fourth-grade students; fourth-grade students are the most compatible audience for the *TTT* program since New York history is taught in the fourth grade. To further limit the variability and strengthen the reliability of the research, the sample included students that are similar in demographics and socioeconomic characteristics (see the next section, “School Selection,” for further details). Additionally, data were collected from both treatment and control groups within consecutive weeks so as to mitigate any differences in student learning that may occur over any lapse in time.

A separate sample of New York City fourth-grade students were selected to participate in the pre-test. RK&A administered the student measures to test the readability and accuracy of the instruments. These students were similar in demographics and socioeconomics to the treatment and control students to bolster the reliability of the pre-test.

SCHOOL SELECTION

STUDY SCHOOLS

During September 2009, RK&A and MCNY used the following criteria to identify a number of potential schools.³

- ◆ New York City schools within one mile of MCNY;
- ◆ Large student body with at least three fourth-grade classes to ensure an adequate sample size;
- ◆ Less than 50 percent English Language Learner population;
- ◆ Similar demographic and socioeconomic profile;
- ◆ Similar test scores on the New York City English Language Arts and Mathematics tests; and,
- ◆ Ideally, the schools would have had a previous relationship with MCNY.⁴

MCNY sent invitations to the principals at three schools that met the above criteria—P.S. 72, P.S. 108, and P.S. 171;⁵ invitations outlined the study design, importance of the research, as well as ensured the confidentiality of the data collected. The invitation also outlined the requirements of the study, which were to:

- ◆ Designate a coordinator/point of contact for MCNY and RK&A at the school;
- ◆ Distribute letters of agreement to participating teachers;
- ◆ Provide class lists for each of the participating classes;
- ◆ Agree to use pre- and post-visit *TTT* materials only as directed by MCNY;
- ◆ Agree that fourth-grade students will NOT participate in any MCNY gallery programs at any time other than specified by this study until after data collection is complete;

³ Using data from the Web site www.greatschools.net, schools were examined by location, free-lunch eligibility, academics, test scores, teacher characteristics, and students' gender, ethnicity, and English language proficiency.

⁴ This item was a preference and not a requirement. It assumes that schools would be more cooperative and receptive to the study; for instance, we know that these schools are inclined to participate in field trip programs.

⁵ We selected three schools to meet our sampling quota of 150 students; P.S. 108 and P.S. 171 have three eligible classrooms with moderate class sizes, which we designated as treatment schools, and P.S. 72 has four eligible classrooms with moderate to high class sizes, which we designated as the control school.

- ◆ Agree that participating teachers cannot attend professional development at MCNY that may compromise the study (i.e., those that address the subject matter taught in *TTT*);
- ◆ Distribute parental permission forms to students in participating classrooms;
- ◆ Schedule participating classrooms to attend *TTT* in February or March 2010 if a treatment school and April or May 2010 if a control school; and,
- ◆ Provide access to students and teachers to administer the instruments.

As an incentive, MCNY provided *TTT* to participating classrooms for free and extended this benefit to *every* fourth-grade class in the participating school. MCNY also made their pre- and post-visit materials available to every fourth-grade teacher in the school, and offered each fourth-grade teacher a free professional development workshop of their choice—after-school or weekend workshops—during the 2009-2010 school year (select workshops were to be available only *after* participation in the study).

Representatives from the three invited schools met with Franny Kent, Director of the Frederick A.O. Schwarz Children's Center at MCNY, and in some cases, with two RK&A Associates to discuss the study and provide a forum for questions.

HUMAN SUBJECT PROTECTION

As required by the U.S. Department of Education, RK&A has secured Federalwide Assurance (FWA) to conduct research with human subjects from the U.S. Office for Human Research Protections. RK&A's FWA number is FWA00007535.

Parental consent was secured for all students participating in the instrument pre-test and study. Permission letters were sent home to parents/guardians and were signed and returned to the teachers in a sealed envelope. MCNY retrieved the forms from the schools. Two copies of the letter were sent home with each child, so that parents could retain one copy for their records. Consent forms were provided in English and Spanish (see Appendix B for the parent/guardian consent forms).

All data generated from the study are confidential. Students' and teachers' names were stricken from all data and replaced with ID numbers. RK&A generated and maintains the ID numbers.

METHODOLOGY AND ANALYSIS

RK&A collected all data onsite. See Table 1 for the data collection schedule.

TABLE I
METHODOLOGICAL TIMELINE

SCHOOL	TASK	DATE
PS 72 (Control)	Data collection	Week of February 22
PS 108 (Treatment)	Fieldtrip	Week of March 1
	Data collection	Week of March 8
PS 171 (Treatment)	Fieldtrip	Week of February 22
	Data collection	Week of March 1

RK&A developed unique and specific instruments to gather data. Methodologies and their corresponding instruments are discussed below.

STUDENT QUESTIONNAIRES

METHODOLOGY

Standardized questionnaires were used to measure attitudes in both treatment and control groups. Questionnaires were selected because standardized information can be easily collected from a large sample of students. Furthermore, data collected through the questionnaire can be compared using various statistical analyses.

Standardized questionnaires were administered by data collectors during school hours. Data collectors verbally administered the questionnaire to each class as a whole.⁶ The data collector read aloud each question and potential responses while students read along and completed their personal (hardcopy) questionnaire (see Appendix C). Students placed their name on the cover page of the questionnaire for identification purposes. After the questionnaires were collected, however, the data collector removed the cover page and identified the questionnaire with a pre-assigned student identification number. Data collectors discarded any questionnaires completed by students whose parents did not grant parental permission.

ANALYSIS

The data were analyzed using SPSS 12.0.1 for Windows, a statistical package for personal computers. Analyses included both descriptive and inferential methods. A 0.05 level of significance was used to preclude findings of little practical significance.⁷ See Appendix D for a listing of all statistical analyses that were run.

DESCRIPTIVE STATISTICS

Frequency distributions were calculated for all categorical variables (e.g., gender, treatment/control group). Summary statistics, including the median (50th percentile), mean (average) and standard deviation (spread of scores: “±” in tables), were calculated for variables measured at an interval level or higher (e.g., ratings of attitudes about museums about history).

INFERENTIAL STATISTICS

To examine the relationship between two categorical variables, cross-tabulation tables were computed to show the joint frequency distribution of the variables, and the chi-square statistic (X^2) was used to test the significance of the relationship. For example, phrases used to describe an artifact were compared according to control and treatment group to determine if the two groups differ with respect to their understanding of that word.

To test for differences in the means of two or more groups, an analysis of variance (ANOVA) was performed and the F-statistic was used to test the significance of the difference. For example, ratings of attitudes about museums about history were compared by gender to determine if attitudes about school differ in boys and girls.

⁶ RK&A administered the questionnaire to all students—those with and without parental consent—but disposed of data collected from students without consent.

⁷ When the level of significance is set to $p = 0.05$, any finding that exists at a probability (p -value) ≤ 0.05 is “significant.” When a finding (such as a relationship between two variables or a difference in rating scores) has a p -value of 0.01, there is a 95 percent probability that the finding exists; that is, 95 out of 100 times, the finding is correct. Conversely, there is a 5 percent probability that the finding would not exist; in other words, 5 out of 100 times, the finding appears by chance.

STUDENT INTERVIEWS

METHODOLOGY

In-depth interviews encourage and motivate interviewees to express their opinions, understandings, and the meaning they construct using language and words that they would naturally use to express themselves (as opposed to the language of the evaluator or researcher).

RK&A interviewed students about a set of pictures related to early New York City history and an object from the time of New Amsterdam (see interview guide in Appendix E); the pictures and object were different from those used in the program. Students were pulled from class two to three at a time; one student was assigned to each data collector at any given time. All data were collected during school hours; each school helped arrange a semi-private space for the data collectors to work. Interviews were audio recorded to facilitate analysis.

RUBRIC

RK&A developed a scoring rubric—a set of criteria linked to learning objectives that is used to assess performance of knowledge, skills, etc. on a continuum—to measure the interviews. Scoring rubrics are useful because they allow qualitative data to be measured in a quantitative way, thus allowing outcomes to be measured. For this study, a rubric was used to measure students' skills related to historical knowledge, historical inquiry, historical perspective, and historical reasoning. For each item, interviews were scored on the scale from 1, "Below Beginning," to 4, "Accomplished." The scoring rubric was developed based on the patterns and trends that emerged from the interview data, along with input from MCNY staff. See Appendix F for the final scoring rubric.

ANALYSIS

Two trained data collectors who did not collect the data and who were not privy to the research hypotheses scored the interviews. One data collector scored all of the data, while the other data collector scored 25 percent of the data. To ensure the reliability of the scoring, RK&A tested the inter-rater reliability of the scores (i.e., compared the data collectors scores), which ranged from 68 percent to 100 percent for each rubric. Data were entered into a computer and analyzed statistically using SPSS 12.0.1. A standard 0.05 level of significance was used to preclude relationships bearing little or no practical significance. See Appendix G for a listing of all statistical analyses that were run.

DESCRIPTIVE STATISTICS

Means were calculated to show a summary of treatment students' achievement.

INFERENTIAL STATISTICS

To examine the relationship between two categorical variables, cross-tabulation tables were computed to show the joint frequency distribution of the variables, and the chi-square statistic (X^2) was used to test the significance of the relationship. For example, scores for Rubric 1a were compared by control and treatment group to determine if the two groups differ with respect to this item.

TEACHER QUESTIONNAIRES

METHODOLOGY

RK&A administered questionnaires to teachers between February and March in conjunction with student data collection at the teachers' school (see Appendix H for the teacher questionnaire). The questionnaires were used to contextualize the findings.

ANALYSIS

Because the sample size is small, RK&A reports the response for each questionnaire item; see Appendix I for a presentation of findings.

REPORTING METHOD

This volume—*Volume I: Report*—describes the research design of the study and presents major findings from all methodologies. Quantitative data are reported in tables and figures along with explanatory text. *Volume 2: Appendix* contains all of the instruments used in the study, a descriptive list of statistical analyses conducted, additional data not included in the report, as well as other miscellaneous information that explains the nuances and finer details of the study.

RELIABILITY AND VALIDITY OF THE STUDY

In designing the study and conducting the research, RK&A has been very mindful of the importance of reliable and valid data and analysis. Some precautions taken to ensure the quality of the data are discussed below.

RESEARCH DESIGN

RK&A understands that examining students' experiences of a particular program is complex. Many factors in a students' life can affect their behavior. To account for the multiple variables that influence student experiences, RK&A carefully structured the instruments and analyses to test for the multiple variables that may account for differences in students' performance.

INSTRUMENT RELIABILITY AND VALIDITY

RK&A developed all instruments according to stringent construction techniques, assuring appropriate item wording, order, format, and internal consistency. All instruments were reviewed and approved by the MCNY staff.

The pre-test of the student questionnaires and interviews ensured the readability and coherence of the instruments. Outcomes of the pre-test were used to further refine the instruments. Additionally, the interview pre-test produced data that were used to refine the development of the scoring rubric.

DATA COLLECTION AND RESPONSE SCORING

Quality data collection is as important as quality research design. To ensure that data collection occurred in an unbiased manner, RK&A hired data collectors who do not know the research hypotheses. RK&A sought data collectors who are graduate students with educational research experiences or individuals with comparable research experiences. RK&A extensively trained data collectors and carefully monitored data collection.

SECTIONS OF THE REPORT:

1. Principal Findings: Student Questionnaires
2. Principal Findings: Student Interviews

PRINCIPAL FINDINGS: STUDENT QUESTIONNAIRES

INTRODUCTION

A total of 154 fourth-grade students from three schools completed the student questionnaire (see Appendix C). Questionnaire findings describe students' attitudes about history, thoughts about history knowledge, attitudes about museums about history, and thoughts about artifacts. For treatment group students, findings also describe students' evaluation of *Traveling Through Time*. Findings are reported by treatment and control group, and statistically significant differences between treatment and control students are indicated.⁸

CONTROL AND TREATMENT GROUPS

Table 2 gives the breakdown for the number of students by school and identifies control and treatment schools. There are 78 control students and 76 treatment students.

TABLE 2
STUDENTS AND SCHOOLS BY YEAR

SCHOOL (GROUP)	STUDENT <i>n</i>
PS 72 (Control)	78
PS 108 (Treatment)	51
PS 171 (Treatment)	25
GRAND TOTAL	154

STUDENT CHARACTERISTICS

This section of the report describes student demographics, including gender, age, language(s) spoken at home, and MCNY visits.

AGE AND GENDER

Table 3 (next page) shows students' ages. Almost all students were either 9 years (67 percent) or 10 years of age (31 percent). Control and treatment students do not differ by age.

⁸ Differences by gender are reported in Appendix J.

TABLE 3
AGE BY GROUP

	GROUP		
	CONTROL	TREATMENT	TOTAL
AGE (n = 154)	%	%	%
8 years	1	0	1
9 years	74	63	69
10 years	23	34	29
11 years	1	3	2
SUMMARY STATISTICS			
Median age	9	9	9
Mean age	9.2	9.4	9.3
Standard deviation	± .49	± .54	± .52

As Table 4 shows, the sample was split evenly between girls and boys (50 percent vs. 50 percent). Control and treatment students do not differ by gender.

TABLE 4
GENDER BY GROUP

	GROUP		
	CONTROL	TREATMENT	TOTAL
GENDER (n = 154)	%	%	%
Boy	53	47	50
Girl	47	53	50

LANGUAGES SPOKEN AT HOME

The questionnaire asked students to identify the language(s) spoken at home (see Table 5). Almost all students said they speak English at home (95 percent), and many also speak Spanish at home (70 percent). Control and treatment group students do not differ by language.

TABLE 5
LANGUAGE(S) SPOKEN AT HOME BY GROUP

	GROUP		
	CONTROL	TREATMENT	TOTAL
LANGUAGE(S) SPOKEN AT HOME (n = 154)	% ¹	% ¹	% ¹
English	92	97	95
Spanish	74	66	70
Other language ²	8	13	10

¹ Column totals exceed 100 percent because some students reported speaking more than one language at home.

Table 6 presents the “other” languages listed by students. Of the many languages listed, Bangla ($n = 6$) and French ($n = 3$) were most frequently identified (see Table 6).

TABLE 6
“OTHER” LANGUAGES SPOKEN AT HOME BY GROUP

“OTHER” LANGUAGE(S) ($n = 16$ STUDENTS)	GROUP		
	CONTROL	TREATMENT	TOTAL ¹
	FREQUENCY	FREQUENCY	FREQUENCY
Bangla/Bengali	2	3	5
French	0	3	3
German	0	2	2
Italian	1	1	2
Arabic	1	0	1
Chinese	1	0	1
Guyanese	0	1	1
Japanese	1	0	1
Sign language	0	1	1

¹One student reported more than one “other” language spoken at home.

MCNY MUSEUM VISITS

All treatment students and almost one-quarter of control students reported visiting MCNY with school sometime over their school career (for treatment students this includes the intended *Traveling Through Time* fieldtrip; 100 percent treatment and 22 percent control) (see Table 7); this is a statistically significant difference. Additionally, almost one-third of treatment students and one-fifth of control students reported visiting MCNY with their family (30 percent treatment and 21 percent control).

TABLE 7
MCNY VISITS BY GROUP

MCNY MUSEUM VISITS	n	GROUP		
		CONTROL	TREATMENT	TOTAL
		%	%	%
Has ever visited with school ¹	154	22	100	60
Has ever visited with family	154	21	30	25

¹ $\chi^2 = 98.421; p = .000$

HISTORY ATTITUDES AND KNOWLEDGE

ATTITUDES ABOUT HISTORY

Students responded to five statements about history on the scale “Strongly disagree – Somewhat disagree – Somewhat agree – Strongly agree.” Depending on the statement, the most favorable response was either “Strongly disagree” or “Strongly agree.” Therefore, to analyze the statements all together, each statement was scored from 1 – 4 points with 1 point given to the least favorable response

and 4 points given to the most favorable response (most favorable response indicated in parentheses following each statement in the table).

Table 8 shows the results by control and treatment groups. Students responded most favorably to the statement, “Learning about history is exciting” (total mean = 3.2), and students responded least favorably to the statement, “Learning about history helps me understand more about myself” (total mean = 2.5).

TABLE 8
ATTITUDES ABOUT HISTORY BY GROUP

LEAST FAVORABLE RESPONSE = 1 MOST FAVORABLE RESPONSE = 4	n	GROUP		
		CONTROL	TREATMENT	TOTAL
		MEAN	MEAN	MEAN
Learning about history is exciting (most favorable response = strongly agree).	154	3.2	3.2	3.2
Studying history is like being a detective (most favorable response = strongly agree).	154	3.0	3.1	3.0
Learning history is only important for school (most favorable response = strongly disagree). ¹	152	2.8	3.2	3.0
I learn about history by talking to my parents/grandparents (most favorable response = strongly agree).	153	2.9	2.9	2.9
Learning about history helps me understand more about myself (most favorable response = strongly agree).	151	2.4	2.6	2.5

¹F = 7.319; p = .008

STATISTICAL DIFFERENCES IN ATTITUDES ABOUT HISTORY BY GROUP

Control and treatment students’ ratings differed in one of the five statements. On the scale 1 (“Strongly agree”) to 4 (“Strongly disagree”):

- ♦ Treatment students are more likely than are control students to strongly disagree with the statement, “Learning history is only important for school” (treatment mean = 3.2 versus control mean = 2.8) (see Table 8).

THOUGHTS ABOUT HISTORY KNOWLEDGE

Students did a sentence completion exercise in which they selected two responses from among six to complete the sentence, “We know what we know about history from _____.”

Table 9 (next page) shows control and treatment students’ responses. The top two selections were: “teachers” (55 percent of total) and “textbooks” (42 percent of total). All other responses were selected by less than one-third of students: “studying objects and maps from the past” (29 percent of total), “the Internet” (25 percent of total), “old photographs” (24 percent of total), “talking with my parents/grandparents about their life” (24 percent of total).

TABLE 9
THOUGHTS ABOUT HISTORY KNOWLEDGE BY GROUP

"WE KNOW WHAT WE KNOW ABOUT HISTORY FROM _____." (n = 154)	GROUP		
	CONTROL	TREATMENT	TOTAL
	% ¹	% ¹	% ¹
Teachers ²	65	43	55
Textbooks	45	40	42
Studying objects and maps from the past	22	36	29
The Internet	27	24	25
Old photographs ³	15	33	24
Talking with my parents/grandparents about their life	23	25	24

¹Column totals exceed 100 percent because students selected two responses.

² $\chi^2 = 7.490; p = .006$

³ $\chi^2 = 6.466; p = .011$

STATISTICAL DIFFERENCES IN THOUGHTS ABOUT HISTORY KNOWLEDGE BY GROUP

Control and treatment students' selections differed in two of the six responses:

- ◆ Control students are more likely than are treatment students to select "teachers" (65 percent of control versus 43 percent of treatment) (see Table 9).
- ◆ Treatment students are more likely than are control students to select "old photographs" (33 percent of treatment versus 15 percent of control) (see Table 9).

ATTITUDES ABOUT HISTORY MUSEUMS

Students responded to four statements about history museums on the scale "Strongly disagree – Somewhat disagree – Somewhat agree – Strongly agree." Depending on the statement, the most favorable response was either "Strongly disagree" or "Strongly agree." Therefore, to analyze the statements all together, each statement was scored from 1 – 4 points with 1 point given to the least favorable response and 4 points given to the most favorable response (most favorable response indicated in parentheses following each statement in the table).

Table 10 (next page) shows the results by control and treatment group. Students responded most favorably to the statement, "The best thing about museums about history is the interesting objects" (total mean = 3.4). Students responded less favorably to the other statements: "Museums about history are too quiet" (total mean = 2.5), "The things I learn about history in school are different from the things I learn about history in museums" (total mean = 2.4), and "In museums about history, I am not allowed to touch anything" (total mean = 2.2).

TABLE 10
ATTITUDES ABOUT HISTORY MUSEUMS BY GROUP

LEAST FAVORABLE RESPONSE = 1 MOST FAVORABLE RESPONSE = 4	n	GROUP		
		CONTROL	TREATMENT	TOTAL
		MEAN	MEAN	MEAN
The best thing about museums about history is the interesting objects (most favorable response = strongly agree). ¹	154	3.6	3.3	3.4
Museums about history are too quiet (most favorable response = strongly disagree).	153	2.4	2.6	2.5
The things I learn about history in school are different from the things I learn about history in museums (most favorable response = strongly disagree).	154	2.3	2.5	2.4
In museums about history, I am not allowed to touch anything (most favorable response = strongly disagree). ²	153	1.8	2.6	2.2

¹F = 7.891; p = .006
²F = 24.691; p = .000

STATISTICAL DIFFERENCES IN ATTITUDES ABOUT HISTORY MUSEUMS BY GROUP

Control and treatment students’ ratings differed in two of the four statements:

- ♦ On the scale 1 (“Strongly disagree”) to 4 (“Strongly agree”), control students are more likely than are treatment students to strongly agree with the statement, “The best thing about museums about history is the interesting object” (control mean = 3.6 versus treatment mean = 3.3) (see Table 10).
- ♦ On the scale 1 (“Strongly agree”) to 4 (“Strongly disagree”), treatment students are more likely than are control students to strongly disagree with the statement, “In museums about history, I am not allowed to touch anything” (treatment mean = 2.6 versus control mean = 1.8) (see Table 10).

THOUGHTS ABOUT ARTIFACTS

Students did a sentence completion exercise in which they selected two responses from among six to complete the sentence, “An artifact is something that _____.”

Table 11 (next page) shows control and treatment students’ responses. The top two selections were: “teaches you about people in the past” (55 percent) and “important people used in the past” (46 percent). The bottom selection was: “my grandparents have” (2 percent).

TABLE 11
THOUGHTS ABOUT ARTIFACTS BY GROUP

"AN ARTIFACT IS SOMETHING THAT _____." (n = 154)	GROUP		TOTAL
	CONTROL	TREATMENT	
	% ¹	% ¹	% ¹
Teaches you about people in the past	50	59	55
Important people used in the past	40	53	46
Is a clue to understand history	40	41	40
Is displayed in museums	37	33	35
You cannot touch ²	28	13	21
My grandparents have	3	1	2

¹Column totals exceed 100 percent because students selected two responses.

² $\chi^2 = 5.295; p = .021$

STATISTICAL DIFFERENCES IN THOUGHTS ABOUT ARTIFACTS BY GROUP

Control and treatment students' selections differed in one of the six responses:

- ♦ Control students are more likely than are treatment students to select "you cannot touch" (28 percent of control versus 13 percent of treatment) (see Table 11).

OPINIONS OF MCNY FIELDTRIP

FAVORITE PARTS OF MCNY FIELDTRIP

Treatment students did a sentence completion exercise in which they selected two responses from among six to complete the sentence, "The best part of my visit to the Museum of the City of New York was _____."

Table 12 shows control and treatment students' responses. The top three selections were: "talking about the past" (53 percent), "learning about New York City" (52 percent), and "touching objects used in the past" (49 percent). The bottom selection was: "making a Delft tile" (7 percent).

TABLE 12
FAVORITE PARTS OF MCNY FIELDTRIP BY GROUP

"THE BEST PART OF MY VISIT TO THE MUSEUM OF THE CITY OF NEW YORK WAS _____." (n = 75)	TREATMENT
	% ¹
Talking about the past	53
Learning about New York City	52
Touching objects used in the past	49
Spending time away from school	21
Visiting a new place	16
Making a Delft tile	7

¹Column totals exceed 100 percent because students selected two responses.

OPINIONS OF MCNY FIELDTRIP

Treatment students responded to six statements about their MCNY fieldtrip on the scale “Strongly disagree – Somewhat disagree – Somewhat agree – Strongly agree.” Depending on the statement, the most favorable response was either “Strongly disagree” or “Strongly agree.” Therefore, to analyze the statements all together, each statement was scored from 1 – 4 points with 1 point given to the least favorable response and 4 points given to the most favorable response (most favorable response indicated in parentheses following each statement in the table).

Table 13 shows the results by control and treatment group. Students responded most favorably to the statements, “The Museum of the City of New York is a good place for kids,” “We talked about interesting things at the Museum of the City of New York,” and “I did not learn anything new about New York City’s history at Museum of the City of New York” (treatment mean = 3.4). Students responded least favorably to the statement, “The person who led us through the Museum of the City of New York talked too much” (mean = 2.9).

TABLE 13
OPINIONS OF MCNY FIELDTRIP

LEAST FAVORABLE RESPONSE = 1 MOST FAVORABLE RESPONSE = 4	n	TREATMENT
		MEAN
The Museum of the City of New York is a good place for kids.	75	3.4
We talked about interesting things at the Museum of the City of New York (most favorable response = strongly agree).	75	3.4
I did not learn anything new about New York City’s history at Museum of the City of New York (most favorable response = strongly disagree).	74	3.4
I felt comfortable discussing my ideas at the Museum of the City of New York (most favorable response = strongly agree).	75	3.3
The trip to Museum of the City of New York was special (most favorable response = strongly agree).	75	3.2
The person who led us through the Museum of the City of New York talked too much (most favorable response = strongly disagree).	75	2.9

PRINCIPAL FINDINGS: STUDENT INTERVIEWS

INTRODUCTION

A total of 150 fourth-grade students from three schools participated in an interview. As part of the interview, students were presented three images related to the settling of New York City, six terms related to the settling of New York City, and an object from the New Amsterdam time period. Findings are reported by treatment and control group, and statistically significant differences between treatment and control students are indicated.⁹ Select interview transcripts are included in Appendix L.

DESCRIPTION OF STUDENTS

Table 14 presents the breakdown of participating students by school and treatment/control group. A total of 150 students participated in the interview—76 treatment students and 74 control students.

TABLE 14
STUDENTS BY GROUP

SCHOOL (GROUP)	STUDENT <i>n</i>
PS 72 (Control)	74
PS 108 (Treatment)	49
PS 171 (Treatment)	27
TOTAL	150

INTERVIEW PROTOCOL

The interview focused on four activities. See Appendix E for the interview guide.

ACTIVITY 1: DISCUSSING PICTURE A

Students were first presented with Picture A, an image of the Dutch and Lenape trading (see Figure 1, next page). Interviewers told students, “This picture shows things that happened a long time ago in what is now New York City.” After giving the students time to examine the picture, interviewers asked students to identify who they think the people are, how the people are the same and different, and what the people are doing (additional standardized probes were used to help students talk about the picture).

⁹ Differences by gender are reported in Appendix J. Additionally, the comparison of control and treatment students is represented in figures in the body of the report; Appendix K contains the same data in tables.

FIGURE 1
PICTURE A



ACTIVITY 2: DISCUSSING PICTURE B AND C

Students were presented Picture B, an image of a Lenape Village, and Picture C, an image of New Amsterdam (see Figure 2 and Figure 3, next page). Interviewers told students that these are other pictures of “what New York City looked like long ago.” After giving the students time to examine each picture, first Picture B and then Picture C, interviewers asked students to identify the places, contrast them, and note which picture comes first on a timeline (additional probes were used to help students talk about the picture).

FIGURE 2
PICTURE B

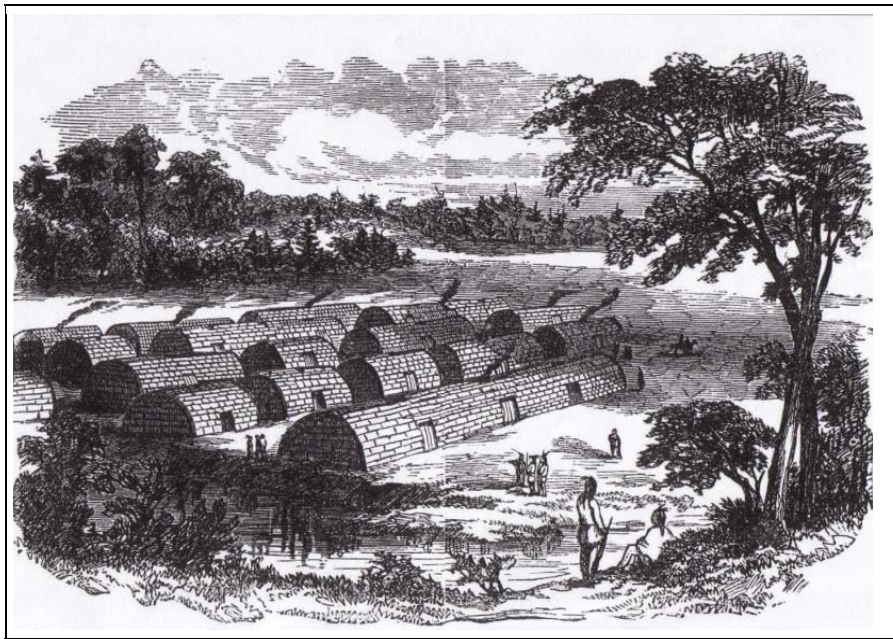
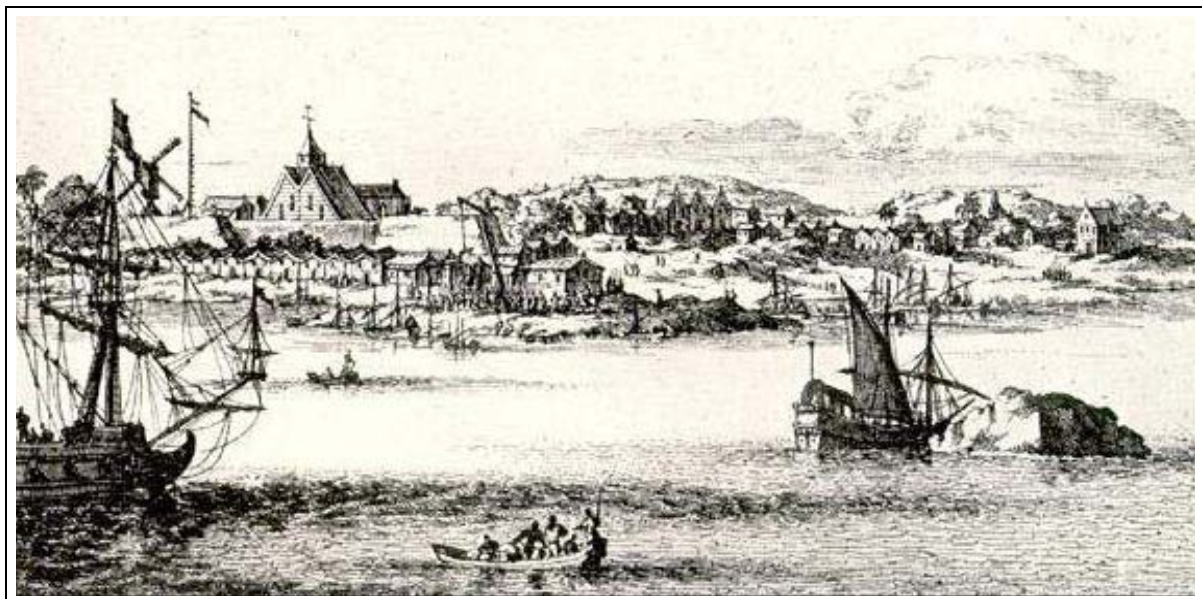


FIGURE 3
PICTURE C



ACTIVITY 3: EXPLAINING THE HISTORY OF NEW YORK CITY WITH IMAGES AND KEY TERMS

Students were presented all three images previously discussed (Pictures A, B, and C) as well as six note cards with one key term on each. Key terms were Lenape, Dutch, New Amsterdam, exploration, beavers, and trade. Interviewers instructed students, “Place these cards on the pictures in a way that explains the history of New York City. You can place more than one card on a picture or one card can touch all three pictures if you want. If you are not sure where to place the cards, take a guess or leave it out.” After giving the students time to place the cards on the pictures, interviewers asked students why they choose to place specific terms on specific pictures and how each term is important to the history of New York City.

ACTIVITY 4: IDENTIFYING A HISTORICAL OBJECT

Lastly, students were presented with a candle maker (see Figure 4, next page). Interviewers told students that “the object is from around the time and place shown in Picture C [New Amsterdam].” Students were encouraged to look at and touch the candle maker. Interviewers then asked students to describe the object and guess what the object may have been used for.

FIGURE 4

CANDLE MAKER



PRESENTATION OF FINDINGS

The findings are organized according to the four skill areas we tested for: historical knowledge, historical perspective, historical inquiry, and historical reasoning.

HISTORICAL KNOWLEDGE is students' factual knowledge about people or events. The three knowledge areas explored in this study are the Lenape, Dutch, and trade between the Lenape and Dutch.

HISTORICAL INQUIRY explores students' ability to explain how they know what they know about an object, illustration, or primary source material. The six inquiry areas explored in this study are the Lenape, Dutch, trade between the Lenape and Dutch, observation of a candle maker, formulating hypotheses for the candle maker, and object identification.

HISTORICAL PERSPECTIVE explores students' understanding of and appreciation for differences among peoples/situations/cultures. The three perspective areas explored in this study are differences between the Lenape and Dutch, difference in belief systems, and cultural variation.

HISTORICAL REASONING explores students' understanding of cause and effect and/or change over time. The eight areas of reasoning assessed in the study were: explaining why the Lenape and Dutch traded with one another, explaining changes in landscape over time, and the significance of the Lenape, Dutch, New Amsterdam, trade, beavers, and exploration to the history of New York City.

HISTORICAL KNOWLEDGE

RUBRIC 1A – HISTORICAL KNOWLEDGE OF THE LENAPE

RUBRIC CRITERIA

Rubric 1a describes the continuum of students’ knowledge of the Lenape based on Picture A (see Table 15).

TABLE 15

CRITERIA FOR RUBRIC 1A – HISTORICAL KNOWLEDGE OF THE LENAPE

1 - Below Beginning	2 – Beginning	3 – Developing	4 – Accomplished
The student describes what the Lenape are wearing (e.g., no clothes, loincloth), but makes no attempt to label them.	The student identifies the Lenape as “Iroquois” or some other incorrect native group.	The student identifies the Lenape as “Native Americans,” “American Indians,” “Indians,” or “Algonquins.”	The student identifies the Lenape as “Lenape.”

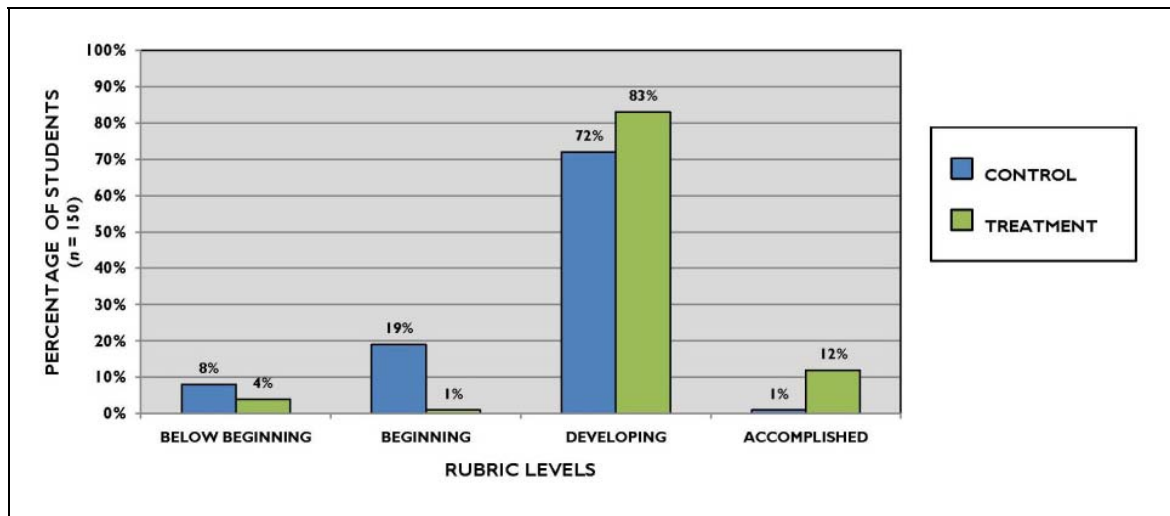
STATISTICAL DIFFERENCES BY GROUP

Figure 5 shows control and treatment students’ achievement on Rubric 1a. Treatment students scored higher than control students:

- ♦ 95 percent of treatment students scored at the “Developing” or “Accomplished” level, while 73 percent of control students scored at the “Developing” or “Accomplished” level.

FIGURE 5

ACHIEVEMENT ON RUBRIC 1A – HISTORICAL KNOWLEDGE OF THE LENAPE



$\chi^2 = 19.506; p = .000$

RUBRIC 1B – HISTORICAL KNOWLEDGE OF THE DUTCH

RUBRIC CRITERIA

Rubric 1b describes the continuum of students’ knowledge of the Dutch based on Picture A (see Table 16).

TABLE 16

CRITERIA FOR RUBRIC 1B – HISTORICAL KNOWLEDGE OF THE DUTCH

1 - Below Beginning	2 – Beginning	3 – Developing	4 – Accomplished
The student describes what the Dutch are wearing (e.g., hats, boots), but makes no attempt to identify them.	The student identifies the Dutch as “explorers,” “settlers,” “westerners,” “pilgrims” or some other generalized term.	The student identifies the Dutch as “Europeans” or as a European group other than the Dutch (e.g., British).	The student identifies the Dutch as “Dutch” or says that they are people from Holland or the Netherlands.

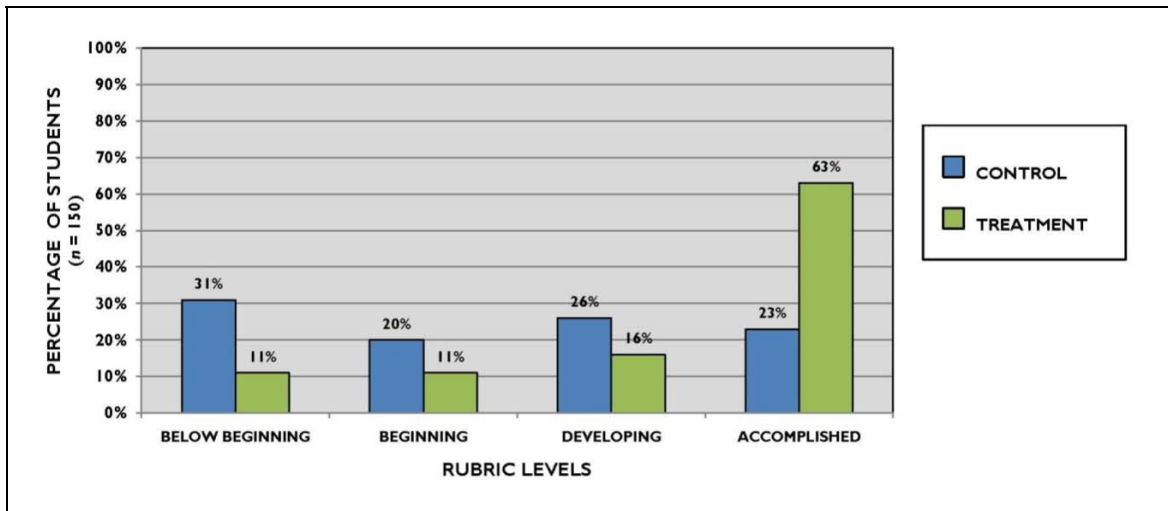
STATISTICAL DIFFERENCES BY GROUP

Figure 6 shows control and treatment students’ achievement on Rubric 1b. Treatment students scored higher than control students:

- ♦ 79 percent of treatment students scored at the “Developing” or “Accomplished” level, while 49 percent of control students scored at the “Developing” or “Accomplished” level.

FIGURE 6

ACHIEVEMENT ON RUBRIC 1B – HISTORICAL KNOWLEDGE OF THE DUTCH



$\chi^2 = 25.732; p = .000$

RUBRIC 1C – HISTORICAL KNOWLEDGE OF TRADE BETWEEN THE LENAPE AND DUTCH

RUBRIC CRITERIA

Rubric 1c describes the continuum of students’ knowledge of trade between the Lenape and the Dutch based on Picture A (see Table 17).

TABLE 17
CRITERIA FOR RUBRIC 1C – HISTORICAL KNOWLEDGE OF TRADE BETWEEN THE LENAPE AND DUTCH

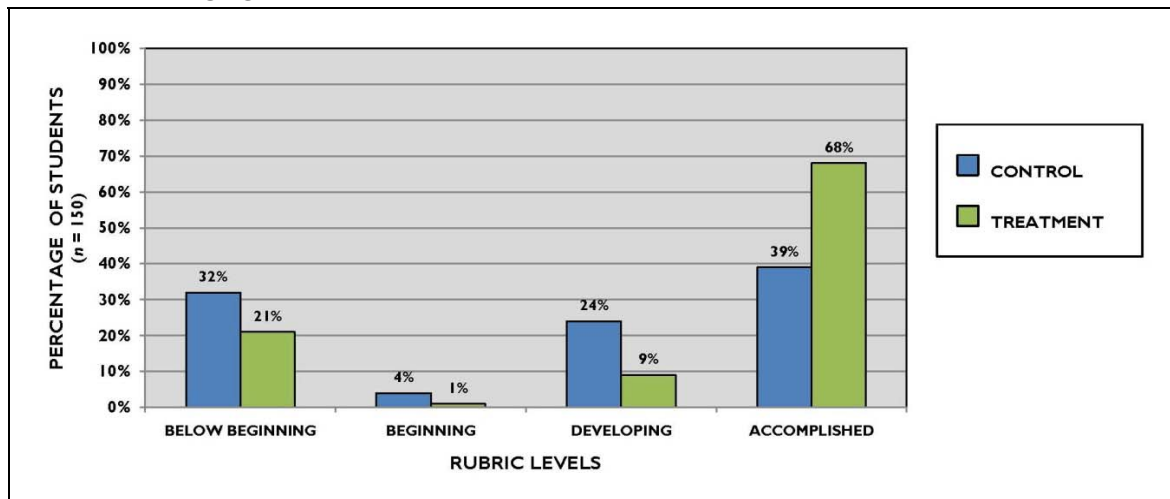
1 - Below Beginning	2 – Beginning	3 – Developing	4 – Accomplished
The student identifies the activity in the image as something other than trading (e.g., working).	The student identifies the activity in the picture <i>primarily</i> as buying or selling materials.	The student identifies the activity in the picture <i>primarily</i> as “trading” and describes how one group gave things to another group—one-sided. OR The student does not use the word “trading” but describes “exchanging” or “giving” items to one another.	The student identifies the activity in the picture <i>primarily</i> as “trading” and describes how both groups exchange materials—reciprocal.

STATISTICAL DIFFERENCES BY GROUP

Figure 7 shows control and treatment students’ achievement on Rubric 1c. Treatment students scored higher than control students:

- ♦ 77 percent of treatment students scored at the “Developing” or “Accomplished” level (68 percent at the “Accomplished” level), while 63 percent of control students scored at the “Developing” or “Accomplished” level (with 39 percent at the “Accomplished” level).

FIGURE 7
ACHIEVEMENT ON RUBRIC 1C – HISTORICAL KNOWLEDGE OF TRADE BETWEEN THE LENAPE AND DUTCH



$\chi^2 = 13.947; p = .003$

HISTORICAL INQUIRY

RUBRIC 2A – HISTORICAL INQUIRY ABOUT THE LENAPE

RUBRIC CRITERIA

Rubric 2a describes the continuum of students’ ability to explain how they know that the people in Picture A are the Lenape (see Table 18).

TABLE 18
CRITERIA FOR RUBRIC 2A – HISTORICAL INQUIRY ABOUT THE LENAPE

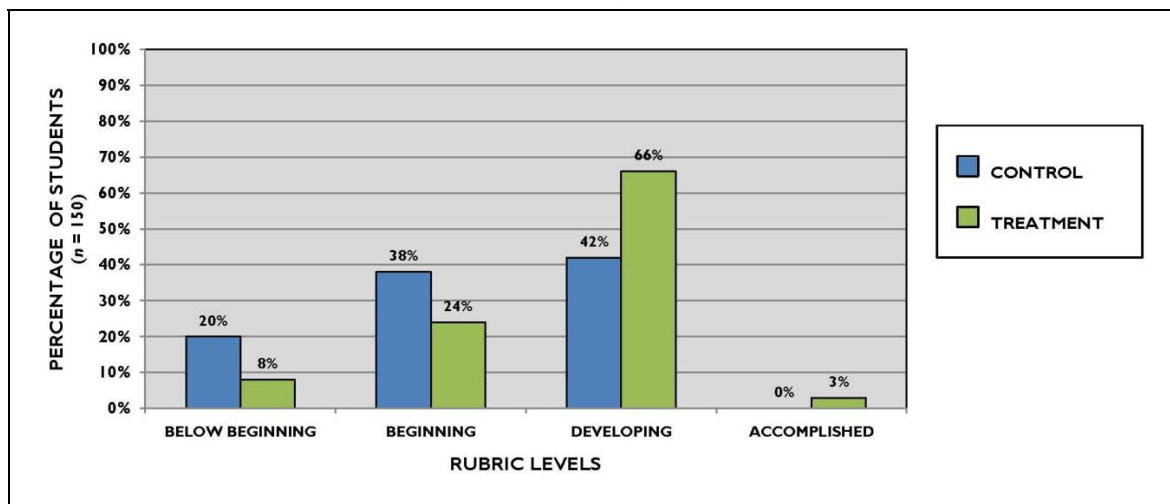
1 - Below Beginning	2 – Beginning	3 – Developing	4 – Accomplished
The student provides completely irrelevant, circular, inaccurate, or no evidence to explain his/her claim about the Lenape.	The student provides partially accurate and relevant evidence to explain his/her claim that the people in the image are Lenape (or that they are who the students says they are). Some of the student’s response is illogical or irrelevant.	The student provides mostly accurate and relevant evidence to explain his/her claim that the people in the image are Lenape (or that they are who the students says they are). The response is vague and/or cursory (hasty and lacking detail though still accurate).	The student provides mostly accurate and relevant evidence to explain his/her claim that the people in the image are Lenape (or that they are who the students says they are.) The response is specific and/or succinct (may be brief but conveys the idea well).

STATISTICAL DIFFERENCES BY GROUP

Figure 8 shows control and treatment students’ achievement on Rubric 2a. Treatment students scored higher than control students:

- ♦ 69 percent of treatment students scored at the “Developing” or “Accomplished” level, while 42 percent of control students scored at the “Developing” or “Accomplished” level.

FIGURE 8
ACHIEVEMENT ON RUBRIC 2A – HISTORICAL INQUIRY ABOUT THE LENAPE



$\chi^2 = 12.463; p = .006$

RUBRIC 2B – HISTORICAL INQUIRY ABOUT THE DUTCH

RUBRIC CRITERIA

Rubric 2b describes the continuum of students’ ability to explain how they know that the people in Picture A are the Dutch (see Table 19).

TABLE 19

CRITERIA FOR RUBRIC 2B – HISTORICAL INQUIRY ABOUT THE DUTCH

1 - Below Beginning	2 – Beginning	3 – Developing	4 – Accomplished
The student provides completely irrelevant, circular, inaccurate, or no evidence to explain his/her claim about the Dutch.	The student provides partially accurate and relevant evidence to explain his/her claim that the people in the image are Dutch (or that they are who the students says they are). Some of the student’s response is illogical or irrelevant.	The student provides mostly accurate and relevant evidence to explain his/her claim that the people in the image are Dutch (or that they are who the students says they are). The response is vague and/or cursory (hasty and lacking detail though still accurate).	The student provides mostly accurate and relevant evidence to explain his/her claim that the people in the image are Dutch (or that they are who the students says they are). The response is specific and/or succinct (may be brief but conveys the idea well).

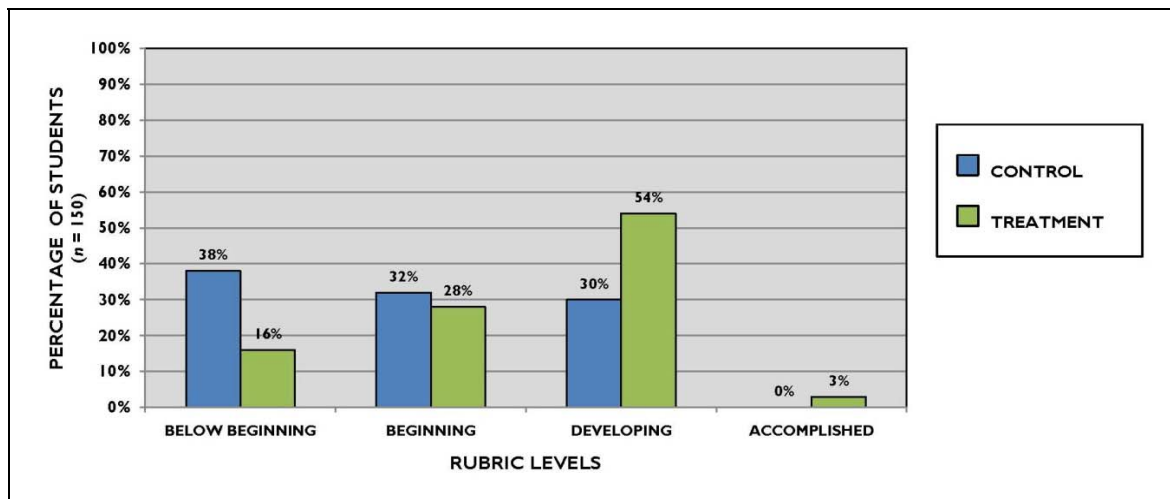
STATISTICAL DIFFERENCES BY GROUP

Figure 9 shows control and treatment students’ achievement on Rubric 2b. Treatment students scored higher than control students:

- ♦ 57 percent of treatment students scored at the “Developing” or “Accomplished” level, while 30 percent of control students scored at the “Developing” level but none at the “Accomplished” level.

FIGURE 9

ACHIEVEMENT ON RUBRIC 2B – HISTORICAL INQUIRY ABOUT THE DUTCH



$\chi^2 = 14.306; p = .003$

RUBRIC 2C – HISTORICAL INQUIRY ABOUT TRADE BETWEEN THE LENAPE AND DUTCH

RUBRIC CRITERIA

Rubric 2c describes the continuum of students’ ability to explain how they know that the people in Picture A are trading (see Table 20).

TABLE 20
CRITERIA FOR RUBRIC 2C - HISTORICAL INQUIRY ABOUT TRADE BETWEEN THE LENAPE AND DUTCH

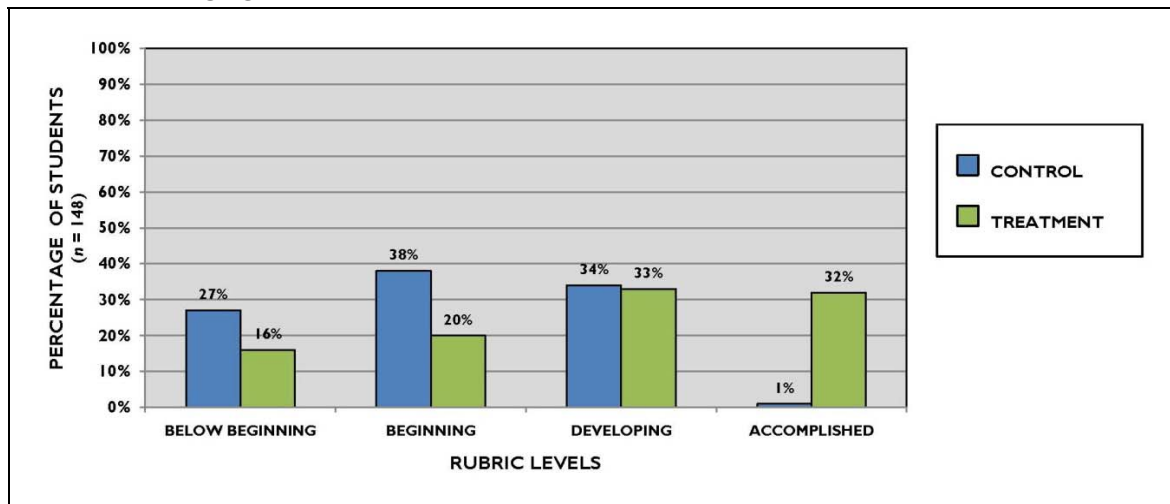
1 - Below Beginning	2 – Beginning	3 – Developing	4 – Accomplished
The student provides completely irrelevant, circular, inaccurate, or no evidence to explain his/her claim about the activity in the image.	The student provides partially accurate and relevant evidence to explain his/her claim that the people in the image are trading (or doing whatever the student says they are doing). Some of the student’s response is illogical or irrelevant.	The student provides mostly accurate and relevant evidence to explain his/her claim that the people in the image are trading (or doing whatever the student says they are doing). The response is vague and/or cursory (hasty and lacking detail though still accurate).	The student provides mostly accurate and relevant evidence to explain his/her claim that the people in the image are trading (or doing whatever the student says they are doing). The response is specific and/or succinct (may be brief but conveys the idea well).

STATISTICAL DIFFERENCES BY GROUP

Figure 10 shows control and treatment students’ achievement on Rubric 2c. Treatment students scored higher than control students:

- ♦ 65 percent of treatment students scored at the “Developing” or “Accomplished” level (with 32 percent at the “Accomplished” level), while 35 percent of control students scored at the “Developing” or “Accomplished” level (with 1 percent at the “Accomplished” level).

FIGURE 10
ACHIEVEMENT ON RUBRIC 2C – HISTORICAL INQUIRY ABOUT TRADE BETWEEN THE LENAPE AND DUTCH



$\chi^2 = 27.068; p = .000$

RUBRIC 2D – HISTORICAL INQUIRY FOR OBSERVATION OF CANDLE MAKER¹⁰

RUBRIC CRITERIA

Rubric 2d describes the continuum of students’ ability to describe the candle maker in detail (see Table 21).

TABLE 21

CRITERIA FOR RUBRIC 2D – HISTORICAL INQUIRY FOR OBSERVATION OF THE CANDLE MAKER

1 - Below Beginning	2 – Beginning	3 – Developing	4 – Accomplished
The student provides a <u>completely vague or off-base</u> description of the object.	The student provides a description that is <u>partially specific and partially vague.</u>	The student provides an <u>incomplete but specific (i.e. concrete)</u> description of the object:	The student provides a <u>complete and specific (i.e. concrete)</u> description of the object.
*listener could not envision the object.	*listener would have a mostly vague picture of the object.	*listener would have a partial picture of the object.	*listener would have a complete, clear mental picture of the object.

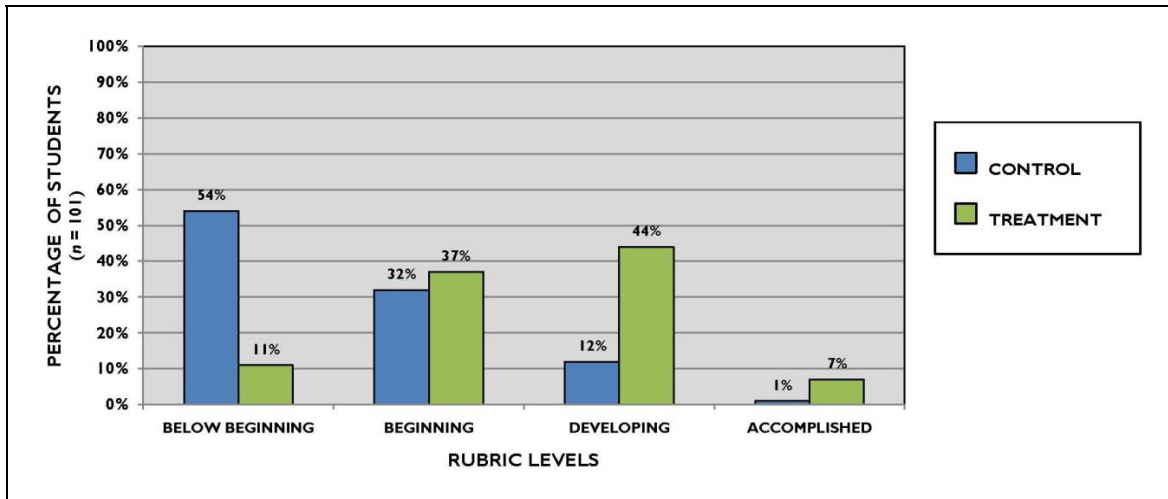
STATISTICAL DIFFERENCES BY GROUP

Figure 11 shows control and treatment students’ achievement on Rubric 2d. Treatment students scored higher than control students:

- ♦ 51 percent of treatment students scored at the “Developing” or “Accomplished” level, while 13 percent of control students scored at the “Developing” or “Accomplished” level.

FIGURE 11

ACHIEVEMENT ON RUBRIC 2D – HISTORICAL INQUIRY FOR OBSERVATION OF THE CANDLE MAKER



$\chi^2 = 21.051; p = .000$

¹⁰ Students from P.S. 108 were not scored on this rubric because we learned from the teacher questionnaire that they had seen candle makers on a fieldtrip to a local historic site.

RUBRIC 2E – HISTORICAL INQUIRY FOR FORMULATING HYPOTHESES ABOUT THE CANDLE MAKER¹¹

RUBRIC CRITERIA

Rubric 2e describes the continuum of students’ ability to hypothesize about what the candle maker is and does (see Table 22).

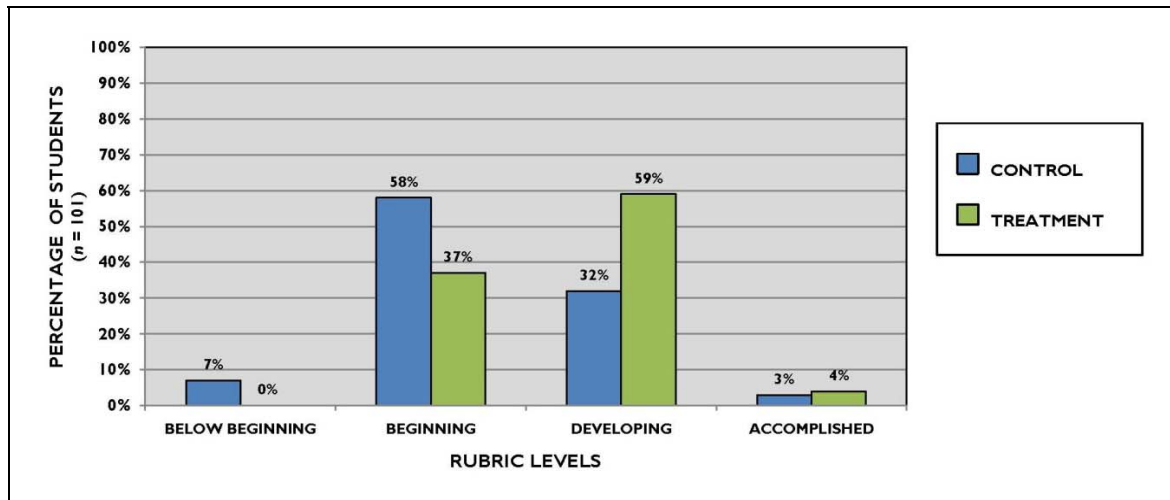
TABLE 22
CRITERIA FOR RUBRIC 2E – HISTORICAL INQUIRY FOR FORMULATING HYPOTHESES ABOUT THE CANDLE MAKER

1 - Below Beginning	2 – Beginning	3 – Developing	4 – Accomplished
The student provides no explanation to identify the object.	The student proposes an explanation to identify the object but provides support that is circular or illogical (repeats the explanation citing it as evidence).	The student proposes a specific explanation to identify the object and provides relevant but vague or cursory support.	The student proposes a specific explanation to identify the object and provides relevant and specific support.

STATISTICAL DIFFERENCES BY GROUP

Figure 12 shows control and treatment students’ achievement on Rubric 2e. There are no statistically significant differences in control and treatment students’ scores.

FIGURE 12
ACHIEVEMENT ON RUBRIC 2E – HISTORICAL INQUIRY FOR FORMULATING HYPOTHESES ABOUT THE CANDLE MAKER



¹¹ Students from P.S. 108 were not scored on this rubric because we learned from the teacher questionnaire that they had seen candle makers on a fieldtrip to a local historic site.

RUBRIC 2F – HISTORICAL INQUIRY FOR OBJECT IDENTIFICATION¹²

RUBRIC CRITERIA

Rubric 2f describes the continuum of students’ ability to correctly identify the object as a candle maker (see Table 23).

TABLE 23

CRITERIA FOR RUBRIC 2F – HISTORICAL INQUIRY FOR OBJECT IDENTIFICATION

1 - Below Beginning	2 – Beginning	3 – Developing	4 – Accomplished
The student does not identify the object.	The student identifies the object but it does not reference candles.	The student says the object is for candles but does not specifically say it is for making candles. May say it is for holding candles.	The student says the object is for making candles.

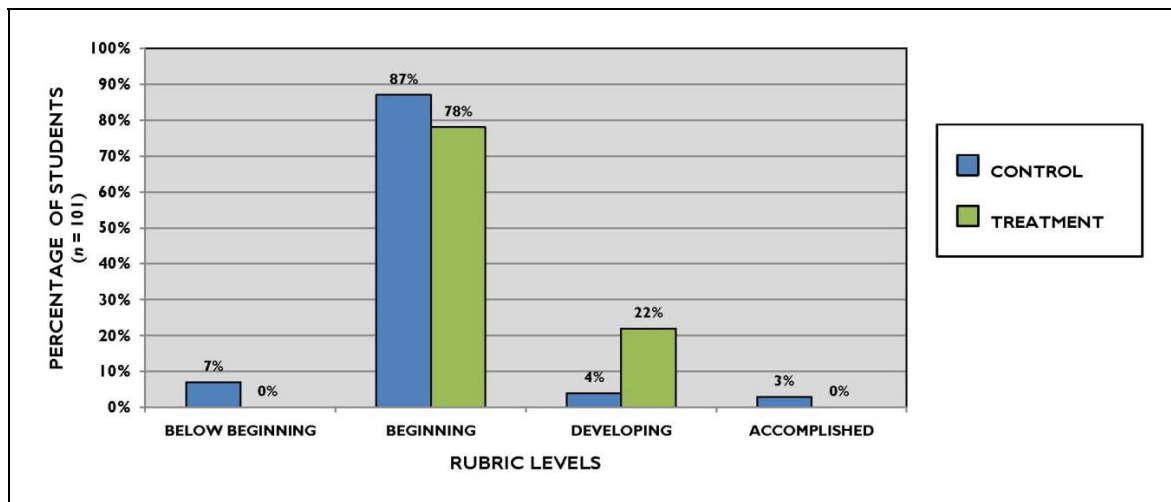
STATISTICAL DIFFERENCES BY GROUP

Figure 13 shows control and treatment students’ achievement on Rubric 2f. Treatment students scored higher than control students:

- ♦ 22 percent of treatment students scored at the “Developing” level but none at the “Accomplished” level, while 7 percent of control students scored at the “Developing” or “Accomplished” level.

FIGURE 13

ACHIEVEMENT ON RUBRIC 2F – FOR HISTORICAL INQUIRY ABOUT OBJECT IDENTIFICATION



$\chi^2 = 10.060; p = .018$

¹² Students from P.S. 108 were not scored on this rubric because we learned from the teacher questionnaire that they had seen candle makers on a fieldtrip to a local historic site.

HISTORICAL PERSPECTIVE

RUBRIC 3A – HISTORICAL PERSPECTIVE OF DIFFERENCES BETWEEN THE LENAPE AND DUTCH

RUBRIC CRITERIA

Rubric 3a describes the continuum of students’ perceptions of the differences between the Lenape and Dutch based on Picture A (see Table 24).

TABLE 24

CRITERIA FOR RUBRIC 3A – HISTORICAL PERSPECTIVE OF DIFFERENCES BETWEEN THE LENAPE AND DUTCH

1 - Below Beginning	2 – Beginning	3 – Developing	4 – Accomplished
<p>The student is unable to name any ways the Lenape and Dutch are different from one another.</p> <p>OR</p> <p>The student provides only inaccurate or illogical example(s) of the ways the Lenape and Dutch are different from one another.</p>	<p>The student primarily talks about the way the Dutch and Lenape are different based on outward appearance as seen in the image (how they are dressed).</p>	<p>The student provides vague, cursory, but <i>mostly</i> accurate example(s) of the ways the Lenape and Dutch are different from one another (i.e., lifestyle difference). Examples go beyond outward appearance as seen in the image.</p>	<p>The student provides thorough, specific, and <i>mostly</i> accurate example(s) of the ways the Lenape and Dutch are different from one another (i.e., lifestyle difference). Examples given go beyond outward appearance as seen in the image.</p>

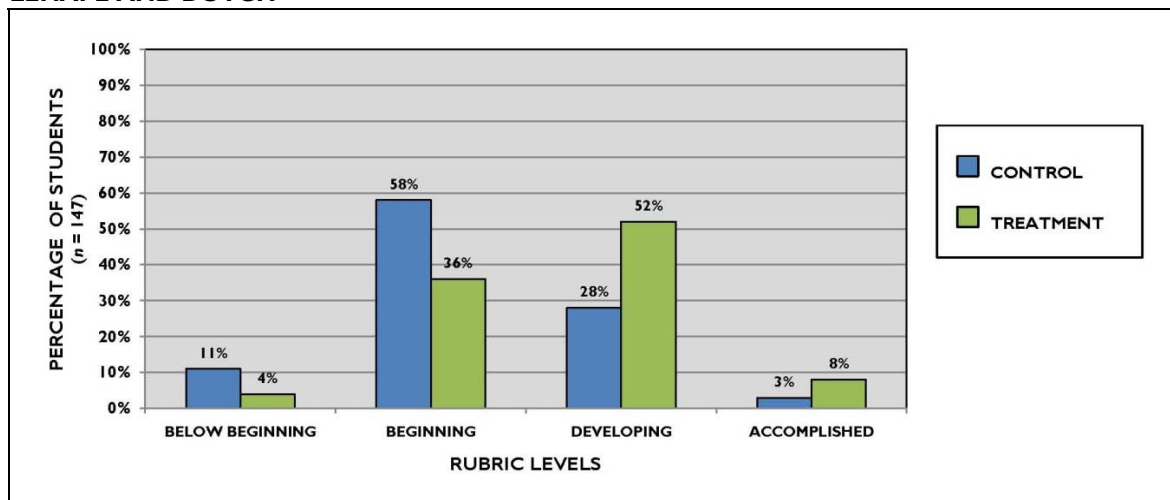
STATISTICAL DIFFERENCES BY GROUP

Figure 14 shows control and treatment students’ achievement on Rubric 3a. Treatment students scored higher than control students:

- ◆ 60 percent of treatment students scored at the “Developing” or “Accomplished” level, while 31 percent of control students scored at the “Developing” or “Accomplished” level.

FIGURE 14

ACHIEVEMENT ON RUBRIC 3A – HISTORICAL PERSPECTIVE OF DIFFERENCES BETWEEN THE LENAPE AND DUTCH



$\chi^2 = 13.597; p = .004$

RUBRIC 3B – HISTORICAL PERSPECTIVE OF DIFFERENT BELIEF SYSTEMS

RUBRIC CRITERIA

Rubric 3b describes the continuum of students’ perceptions of the different belief systems of the Lenape and Dutch based on Picture A, B, or C (see Table 25).

TABLE 25

CRITERIA FOR RUBRIC 3B – HISTORICAL PERSPECTIVE OF DIFFERENT BELIEF SYSTEMS

1 - Below Beginning	2 – Beginning	3 – Developing	4 – Accomplished
The student does not describe differences between the Dutch and Lenape.	The student describes differences between the Dutch and Lenape, but makes no reference to different belief systems, such as their beliefs on land ownership and use of resources.	The student <i>implicitly</i> says that the Dutch and Lenape had different belief or value systems, such as their beliefs on land ownership and use of resources.	The student <i>explicitly</i> says that the Dutch and Lenape had different belief or value systems, such as their beliefs on land ownership and use of resources.

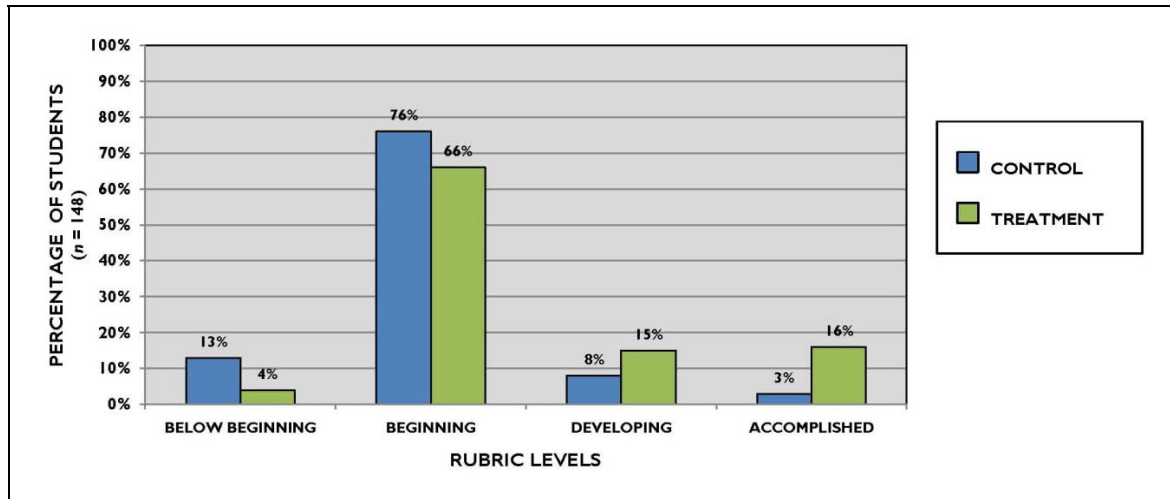
STATISTICAL DIFFERENCES BY GROUP

Figure 15 shows control and treatment students’ achievement on Rubric 3b. Treatment students scored higher than control students:

- ♦ 31 percent of treatment students scored at the “Developing” or “Accomplished” level, while 11 percent of control students scored at the “Developing” or “Accomplished” level.

FIGURE 15

ACHIEVEMENT ON RUBRIC 3B – HISTORICAL PERSPECTIVE OF DIFFERENT BELIEF SYSTEMS



$\chi^2 = 11.752; p = .008$

RUBRIC 3C – HISTORICAL PERSPECTIVE OF CULTURAL VARIATION

RUBRIC CRITERIA

Rubric 3c describes the continuum of students’ acceptance of the differences between the Lenape and the Dutch based on Picture A, B, or C (see Table 26).

TABLE 26

CRITERIA FOR RUBRIC 3C – HISTORICAL PERSPECTIVE OF CULTURAL VARIATION

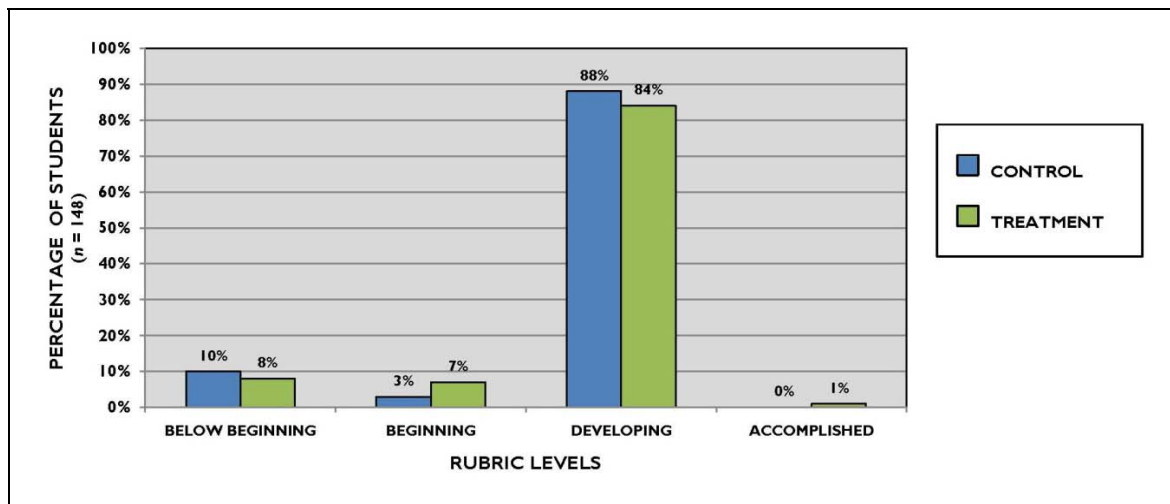
1 - Below Beginning	2 – Beginning	3 – Developing	4 – Accomplished
In describing the different people in the image, the student passes judgment that one group is better than another. He/she states some of the differences as opinions (such as saying that one group is nicer than another). He/she goes further to state erroneous and/or slanderous opinion (such as the Dutch took Lenape as slaves).	In describing the different people in the image, the student passes judgment that one group is better than another. He/she states some of the differences as opinions (such as saying that one group is nicer than another).	In describing the different people in the image, the student does not pass judgment that one group is better than another. He/she states differences as facts rather than opinions.	In discussing the different people in the image, the student does not pass judgment that one group is better than another. He/she also expresses an awareness of this understanding and may say something like, “The Dutch and Lenape were different from one another but neither was better than the other.”

STATISTICAL DIFFERENCES BY GROUP

Figure 16 shows control and treatment students’ achievement on Rubric 3c. There are no statistically significant differences in control and treatment students’ scores.

FIGURE 16

ACHIEVEMENT ON RUBRIC 3C - HISTORICAL PERSPECTIVE OF CULTURAL VARIATION



HISTORICAL REASONING

RUBRIC 4A – HISTORICAL REASONING ABOUT WHY THE LENAPE AND DUTCH TRADED

RUBRIC CRITERIA

Rubric 4a describes the continuum of students’ ability to explain why the people in Picture A wanted to trade (see Table 27).

TABLE 27

CRITERIA FOR RUBRIC 4A - HISTORICAL REASONING ABOUT WHY THE LENAPE AND DUTCH TRADED

1 - Below Beginning	2 – Beginning	3 – Developing	4 – Accomplished
<p>The student does not explain why people in the image traded or exchanged with one another.</p> <p>OR</p> <p>The student does not identify the activity as trading or exchanging goods.</p>	<p>The student is completely inaccurate in identifying the things the people traded (does not name fur, beaver, metal, or tools at all).</p> <p>AND/OR</p> <p>The student provides an illogical or irrelevant explanation for why the people in the image traded or exchanged items with one another.</p>	<p>The student accurately identifies what the people are trading (fur, beaver, metal, tools) and explains that the people in the image traded/ exchanged these items with one another to get things the other did not have but needed. The response is vague and/or cursory (hasty and lacking detail though still accurate). Response may also include a few things that were not actually traded.</p>	<p>The student accurately identifies what the people are trading (fur, beaver, metal, tools) and explains that the people in the image traded/ exchanged these items to get things the other did not have but needed. The response is specific (may be brief but conveys the idea well). Response may also include a few things that were not actually traded.</p>

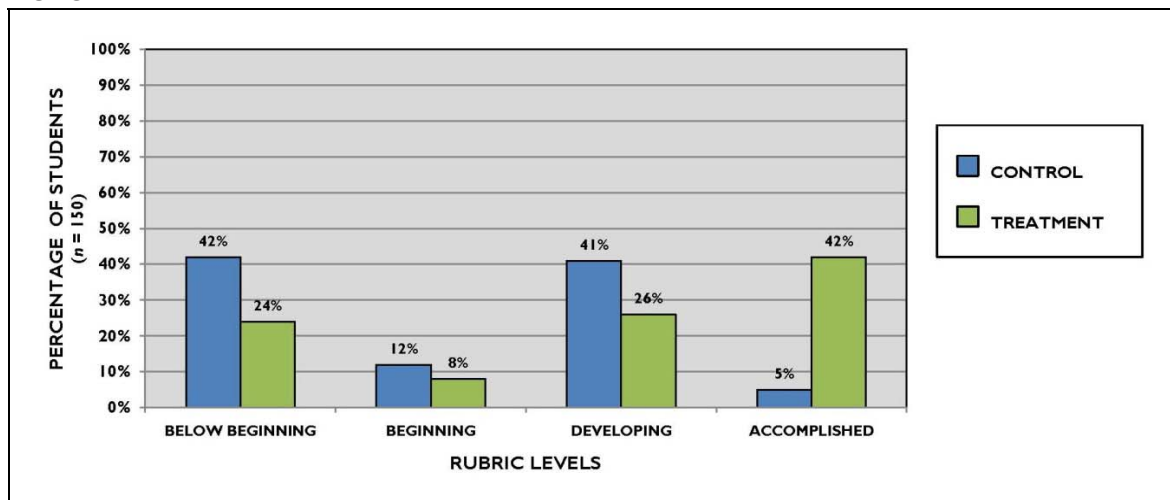
STATISTICAL DIFFERENCES BY GROUP

Figure 17 shows control and treatment students’ achievement on Rubric 4a. Treatment students scored higher than control students:

- ♦ 68 percent of treatment students scored at the “Developing” or “Accomplished” level, while 46 percent of control students scored at the “Developing” or “Accomplished” level.

FIGURE 17

ACHIEVEMENT ON RUBRIC 4A - HISTORICAL REASONING ABOUT WHY THE LENAPE AND DUTCH TRADED



$\chi^2 = 27.805; p = .000$

RUBRIC 4B – HISTORICAL REASONING ABOUT CHANGING LANDSCAPE OF NEW YORK CITY

RUBRIC CRITERIA

Rubric 4b describes the continuum of students’ ability to explain that Picture B and C demonstrate change over time (see Table 28).

TABLE 28

CRITERIA FOR RUBRIC 4B – HISTORICAL REASONING ABOUT CHANGING LANDSCAPE OF NEW YORK CITY

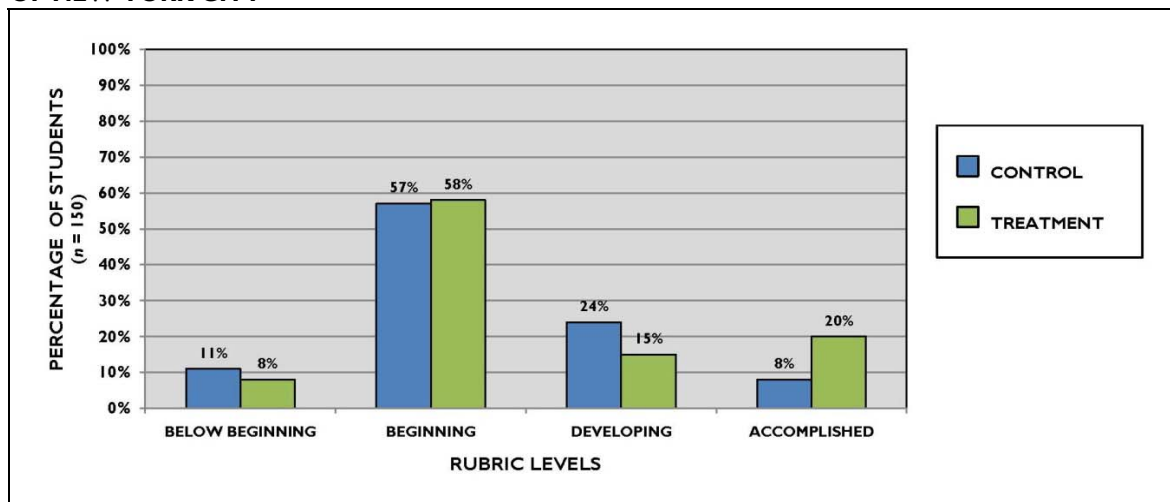
1 - Below Beginning	2 – Beginning	3 – Developing	4 – Accomplished
<p>The student places Picture C first on the timeline.</p> <p>OR</p> <p>The student places Picture B first on the timeline. However, he/she is unable to describe the differences between the two images or describes illogical differences.</p>	<p>The student places Picture B first on the timeline. But the student only <i>implicitly</i> indicates change over time. The student describes visible differences between the two images such as mostly trees to mostly buildings, but does not explicitly indicate knowledge of the relationship between the two images as representing changes over time.</p>	<p>The student places Picture B first on the timeline. Additionally, the student <i>explicitly</i> describes a relationship between the two images as representing changes over time, from mostly untouched natural resources to a built environment.</p>	<p>The student places Picture B first on the timeline. Additionally, the student <i>explicitly</i> describes a relationship between the two images as representing changes over time, from mostly untouched natural resources to a built environment. The student goes further and indicates that these images represent the change from a mostly Native American population to a Dutch settlement in what is now NYC.</p>

STATISTICAL DIFFERENCES BY GROUP

Figure 18 shows control and treatment students’ achievement on Rubric 4b. There are no statistically significant differences in control and treatment students’ scores.

FIGURE 18

ACHIEVEMENT ON RUBRIC 4B – HISTORICAL REASONING ABOUT CHANGING LANDSCAPE OF NEW YORK CITY



RUBRIC 4C – HISTORICAL REASONING ABOUT THE SIGNIFICANCE OF THE LENAPE TO NEW YORK CITY HISTORY

RUBRIC CRITERIA

Rubric 4c describes the continuum of students’ ability to explain the relationship between the Lenape and New York City history based on Picture A, B, or C, and the key terms (see Table 29).

TABLE 29
CRITERIA FOR RUBRIC 4C – HISTORICAL REASONING ABOUT THE SIGNIFICANCE OF THE LENAPE TO NEW YORK CITY HISTORY

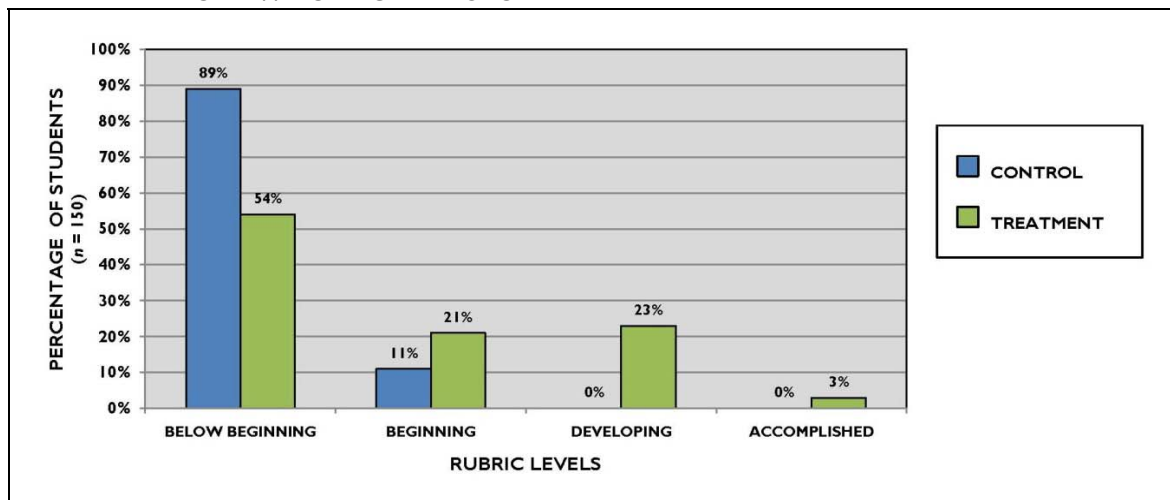
1 - Below Beginning	2 – Beginning	3 – Developing	4 – Accomplished
The student does not explain the relationship of the Lenape to NYC at all or explains it in a way that is completely inaccurate or illogical (seems to be guessing).	The student provides a partially accurate but also somewhat irrelevant explanation to describe the relationship of Lenape to NYC. (Student seems to have some knowledge but is confused or is making an educated guess)	The student provides a somewhat logical and accurate explanation to describe the relationship of Lenape to the history of NYC. The response is vague and/or cursory (hasty and lacking detail though still accurate).	The student provides a mostly logical and accurate explanation to describe the relationship of Lenape to the history of NYC. The response is specific and/or succinct (may be brief but conveys the idea well).

STATISTICAL DIFFERENCES BY GROUP

Figure 19 shows control and treatment students’ achievement on Rubric 4c. Treatment students scored higher than control students:

- ♦ 26 percent of treatment students scored at the “Developing” or “Accomplished” level, while no control students scored at the “Developing” or “Accomplished” level.

FIGURE 19
ACHIEVEMENT ON RUBRIC 4C – HISTORICAL REASONING ABOUT THE SIGNIFICANCE OF THE LENAPE TO NEW YORK CITY HISTORY



$\chi^2 = 26.616; p = .000$

RUBRIC 4D – HISTORICAL REASONING ABOUT THE SIGNIFICANCE OF THE DUTCH IN NEW YORK CITY HISTORY

RUBRIC CRITERIA

Rubric 4d describes the continuum of students’ ability to explain the relationship between the Dutch and New York City history based on Picture A, B, or C, and the key terms (see Table 30).

TABLE 30

CRITERIA FOR RUBRIC 4D – HISTORICAL REASONING ABOUT THE SIGNIFICANCE OF THE DUTCH TO NEW YORK CITY HISTORY

1 - Below Beginning	2 – Beginning	3 – Developing	4 – Accomplished
The student does not explain the relationship of the Dutch to NYC at all or explains it in a way that is completely inaccurate or illogical (seems to be guessing).	The student provides partially accurate but also somewhat irrelevant explanation to describe the relationship of Dutch to NYC. (Student seems to have some knowledge but is confused or is making an educated guess).	The student provides a somewhat logical and accurate explanation to describe the relationship of Dutch to the history of NYC. The response is vague and/or cursory (hasty and lacking detail though still accurate).	The student provides a mostly logical and accurate explanation to describe the relationship of Dutch to the history of NYC. The response is specific and/or succinct (may be brief but conveys the idea well).

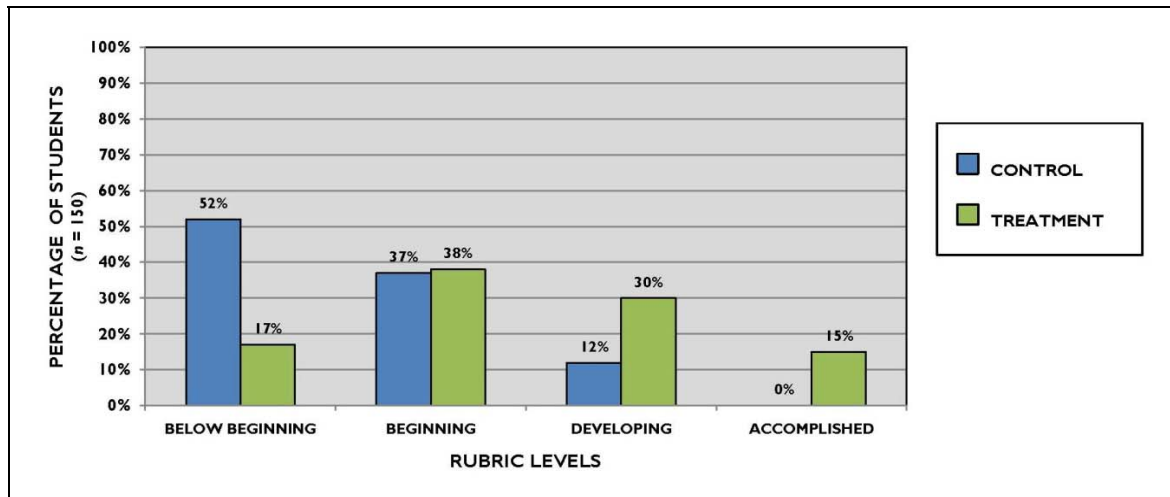
STATISTICAL DIFFERENCES BY GROUP

Figure 20 shows control and treatment students’ achievement on Rubric 4d. Treatment students scored higher than control students:

- ♦ 45 percent of treatment students scored at the “Developing” or “Accomplished” level, while 12 percent of control students scored at the “Developing” level and none scored at the “Accomplished” level.

FIGURE 20

ACHIEVEMENT ON RUBRIC 4D – HISTORICAL REASONING ABOUT THE SIGNIFICANCE OF THE DUTCH TO NEW YORK CITY HISTORY



$\chi^2 = 29.430; p = .000$

RUBRIC 4E – HISTORICAL REASONING ABOUT THE SIGNIFICANCE OF NEW AMSTERDAM TO NEW YORK CITY HISTORY

RUBRIC CRITERIA

Rubric 4e describes the continuum of students’ ability to explain the relationship between New Amsterdam and New York City history based on Picture A, B, or C, and the key terms (see Table 31).

TABLE 31
CRITERIA FOR RUBRIC 4E – HISTORICAL REASONING ABOUT THE SIGNIFICANCE OF NEW AMSTERDAM TO NEW YORK CITY HISTORY

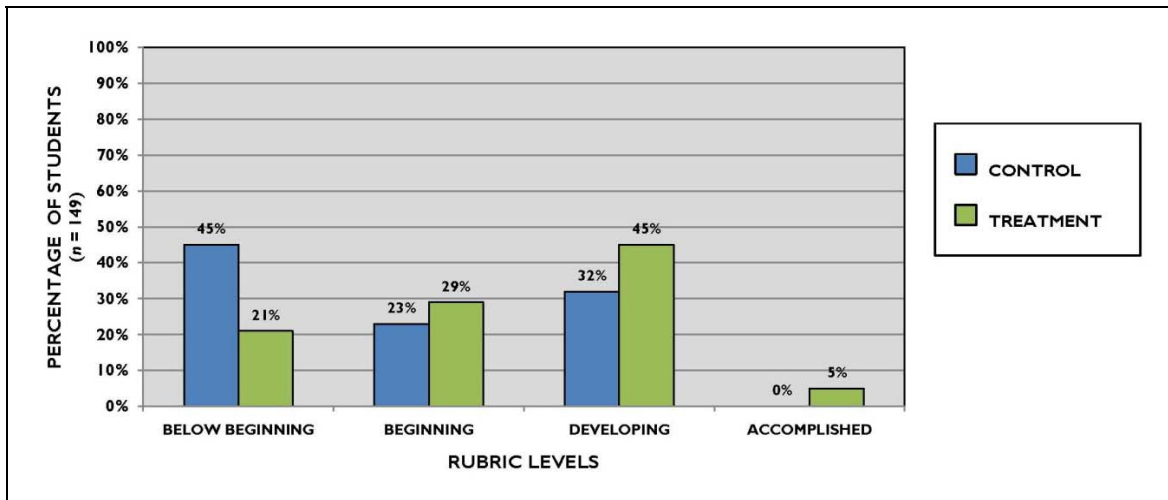
1 - Below Beginning	2 – Beginning	3 – Developing	4 – Accomplished
The student does not explain the relationship of the New Amsterdam to NYC at all or explains it in a way that is completely inaccurate or illogical (seems to be guessing).	The student provides partially accurate but also somewhat irrelevant explanation to describe the relationship of New Amsterdam to NYC. (Student seems to have some knowledge but is confused or is making an educated guess).	The student provides a somewhat logical and accurate explanation to describe the relationship of New Amsterdam to the history of NYC. The response is vague and/or cursory (hasty and lacking detail though still accurate).	The student provides a mostly logical and accurate explanation to describe the relationship of New Amsterdam to the history of NYC. The response is specific and/or succinct (may be brief but conveys the idea well).

STATISTICAL DIFFERENCES BY GROUP

Figure 21 shows control and treatment students’ achievement on Rubric 4e. Treatment students scored higher than control students:

- ♦ 50 percent of treatment students scored at the “Developing” or “Accomplished” level, while 32 percent of control students scored at the “Developing” level and none scored at the “Accomplished” level.

FIGURE 21
ACHIEVEMENT ON RUBRIC 4E – HISTORICAL REASONING ABOUT THE SIGNIFICANCE OF NEW AMSTERDAM TO NEW YORK CITY HISTORY



$\chi^2 = 12.606; p = .006$

RUBRIC 4F – HISTORICAL REASONING ABOUT THE SIGNIFICANCE OF EXPLORATION TO NEW YORK CITY HISTORY

RUBRIC CRITERIA

Rubric 4f describes the continuum of students’ ability to explain the relationship between exploration and New York City history based on Picture A, B, or C, and the key terms (see Table 32).

TABLE 32
CRITERIA FOR RUBRIC 4F – HISTORICAL REASONING ABOUT THE SIGNIFICANCE OF EXPLORATION TO NEW YORK CITY HISTORY

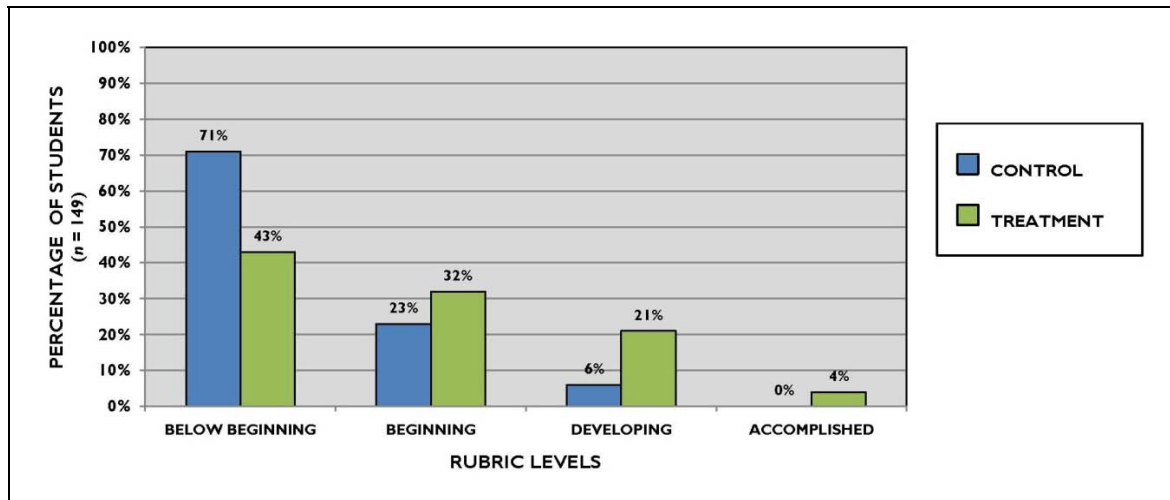
1 - Below Beginning	2 – Beginning	3 – Developing	4 – Accomplished
The student does not explain the relationship of exploration to NYC at all or explains it in a way that is completely inaccurate or illogical (seems to be guessing).	The student provides partially accurate but also somewhat irrelevant explanation to describe the relationship of exploration to NYC. (Student seems to have some knowledge but is confused or is making an educated guess).	The student provides a somewhat logical and accurate explanation to describe the relationship of exploration to the history of NYC. The response is vague and/or cursory (hasty and lacking detail though still accurate).	The student provides a mostly logical and accurate explanation to describe the relationship of exploration to the history of NYC. The response is specific and/or succinct (may be brief but conveys the idea well).

STATISTICAL DIFFERENCES BY GROUP

Figure 22 shows control and treatment students’ achievement on Rubric 4f. Treatment students scored higher than control students:

- ♦ 25 percent of treatment students scored at the “Developing” or “Accomplished” level, while 6 percent of control students scored at the “Developing” level and none scored at the “Accomplished” level.

FIGURE 22
ACHIEVEMENT ON RUBRIC 4F – HISTORICAL REASONING ABOUT THE SIGNIFICANCE OF EXPLORATION TO NEW YORK CITY HISTORY



$\chi^2 = 15.588; p = .001$

RUBRIC 4G – HISTORICAL REASONING ABOUT THE SIGNIFICANCE OF BEAVERS TO NEW YORK CITY HISTORY

RUBRIC CRITERIA

Rubric 4g describes the continuum of students’ ability to explain the relationship between beavers and New York City history based on Picture A, B, or C, and the key terms (see Table 33).

TABLE 33
CRITERIA FOR RUBRIC 4G – HISTORICAL REASONING ABOUT THE SIGNIFICANCE OF BEAVERS TO NEW YORK CITY HISTORY

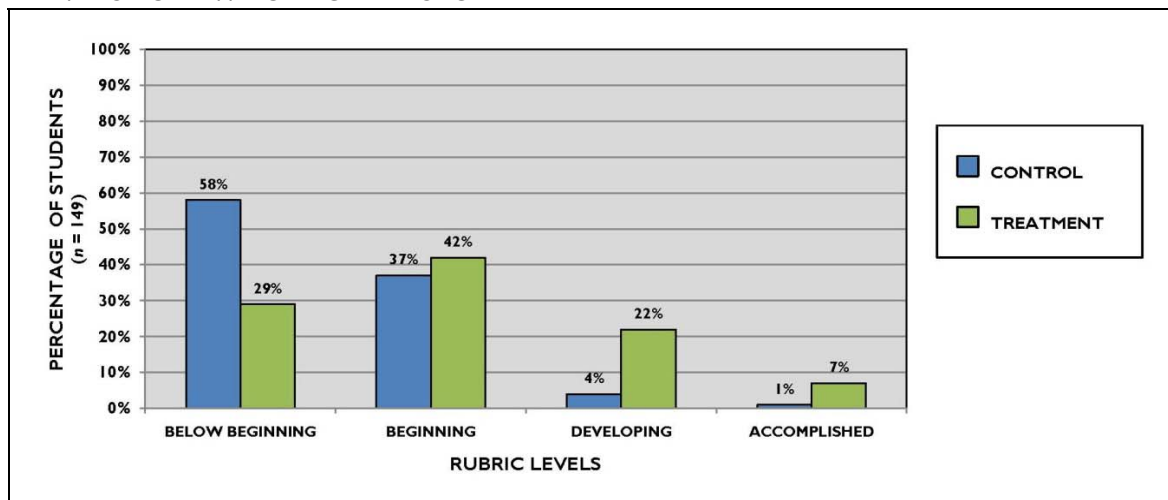
1 - Below Beginning	2 – Beginning	3 – Developing	4 – Accomplished
The student does not explain the relationship of the beavers to NYC at all or explains it in a way that is completely inaccurate or illogical (seems to be guessing).	The student provides partially accurate but also somewhat irrelevant explanation to describe the relationship of beavers to NYC. (Student seems to have some knowledge but is confused or is making an educated guess).	The student provides a somewhat logical and accurate explanation to describe the relationship of beavers to the history of NYC. The response is vague and/or cursory (hasty and lacking detail though still accurate).	The student provides a mostly logical and accurate explanation to describe the relationship of beavers to the history of NYC. The response is specific and/or succinct (may be brief but conveys the idea well).

STATISTICAL DIFFERENCES BY GROUP

Figure 23 shows control and treatment students’ achievement on Rubric 4g. Treatment students scored higher than control students:

- ♦ 29 percent of treatment students scored at the “Developing” or “Accomplished” level, while 5 percent of control students scored at the “Developing” or “Accomplished” level.

FIGURE 23
ACHIEVEMENT ON RUBRIC 4G – HISTORICAL REASONING ABOUT THE SIGNIFICANCE OF BEAVERS TO NEW YORK CITY HISTORY



$\chi^2 = 19.088; p = .000$

RUBRIC 4H – HISTORICAL REASONING ABOUT THE SIGNIFICANCE OF TRADE TO NEW YORK CITY

RUBRIC CRITERIA

Rubric 4h describes the continuum of students’ ability to explain the relationship between trade and New York City history based on Picture A, B, or C, and the key terms (see Table 34).

TABLE 34
CRITERIA FOR RUBRIC 4H – HISTORICAL REASONING ABOUT THE SIGNIFICANCE OF TRADE TO NEW YORK CITY HISTORY

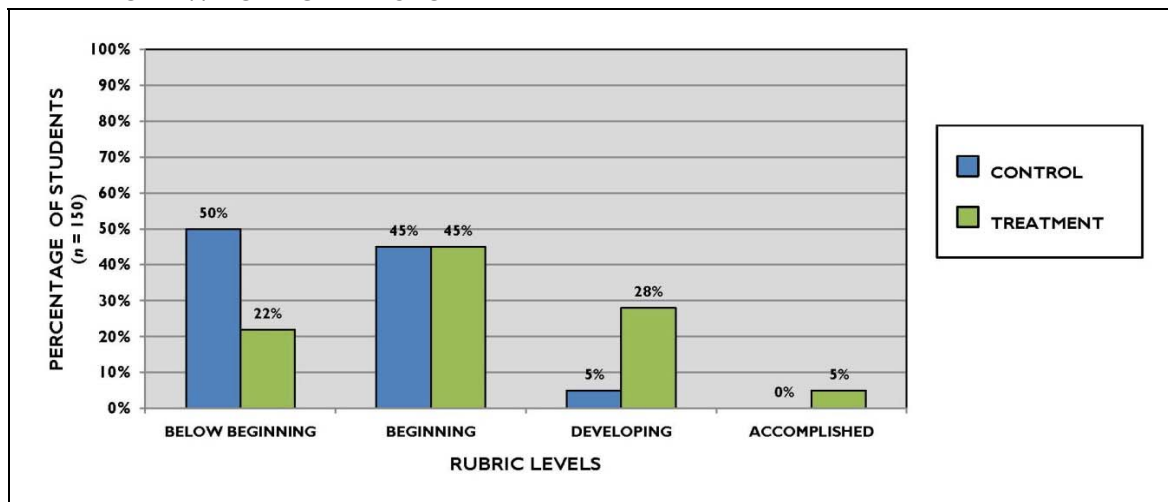
1 - Below Beginning	2 – Beginning	3 – Developing	4 – Accomplished
The student does not explain the relationship of trade to NYC at all or explains it in a way that is completely inaccurate or illogical (seems to be guessing).	The student provides partially accurate but also somewhat irrelevant explanation to describe the relationship of trade to NYC. (Student seems to have some knowledge but is confused or is making an educated guess)	The student provides a somewhat logical and accurate explanation to describe the relationship of trade to the history of NYC. The response is vague and/or cursory (hasty and lacking detail though still accurate).	The student provides a mostly logical and accurate explanation to describe the relationship of trade to the history of NYC. The response is specific and/or succinct (may be brief but conveys the idea well).

STATISTICAL DIFFERENCES BY GROUP

Figure 24 shows control and treatment students’ achievement on Rubric 4h. Treatment students scored higher than control students:

- ♦ 33 percent of treatment students scored at the “Developing” or “Accomplished” level, while 5 percent of control students scored at the “Developing” level and none scored at the “Accomplished” level.

FIGURE 24
ACHIEVEMENT ON RUBRIC 4H – HISTORICAL REASONING ABOUT THE SIGNIFICANCE OF TRADE TO NEW YORK CITY HISTORY



$\chi^2 = 22.960; p = .000$

SUMMARY OF TREATMENT STUDENTS' SCORES

Figure 25 shows a summary of treatment students' mean scores for each rubric. Treatment students scored highest in historical knowledge (top three mean scores) and lowest in historical reasoning (bottom four mean scores).

FIGURE 25
SUMMARY OF TREATMENT STUDENTS' SCORES

