

**The Impact of Authentic Pedagogy on Student Learning in Tenth Grade History Courses**

by

Lamont E. Maddox

A dissertation submitted to the Graduate Faculty of  
Auburn University  
in partial fulfillment of the  
requirements for the Degree of  
Doctor of Philosophy

Auburn, Alabama  
May 7, 2012

Keywords: Assessment, Inquiry, Authentic Intellectual Work

Copyright 2012 by Lamont E. Maddox

Approved by

John W. Saye, Jr., Chair, Professor of Curriculum & Teaching  
Jada Kohlmeier, Associate Professor of Curriculum & Teaching  
Kathryn H. Braund, Professor of History

## Abstract

This mixed-methods study examined the impact of varying levels of authentic pedagogy on student learning in 9<sup>th</sup> and 10<sup>th</sup> grade history classrooms. The sample included four junior high teachers and four high school teachers. During the initial phase of the study, instructional artifacts (tasks) and classroom observational data were collected and analyzed to determine the level of authentic pedagogy students experienced in their classes. Participating teachers were assigned an authentic pedagogy score based on this analysis which was used as the primary independent variable in subsequent statistical analyses designed to evaluate student learning outcomes. The findings suggest that authentic pedagogy has a small, but positive impact on student performance on the Alabama High School Graduation Exam. Classroom level comparisons suggest that students who receive higher levels of authentic pedagogy were not put at a significant disadvantage on a test of lower order knowledge. The study also evaluated the impact of authentic pedagogy on higher order learning outcomes and various subgroups of students (i.e. race, gender, etc.). Due to the small sample of teachers, results should be viewed as extremely tentative and limited to the setting where the study was conducted.

## Acknowledgments

A study as ambitious in scope and duration as this one would never have been possible without the assistance of many people. I first would like to give thanks to God through which all things are possible. Next, I'd like to express my gratitude to the participants of this study, to include the school system itself, for agreeing to support my research. The central office staff went above and beyond the call of duty in providing the necessary coded student data. The teachers were also very gracious for inviting me into their classrooms. The professionalism of everyone involved is to be commended.

I'd also like to thank Dr. John W. Saye, Dr. Jada Kohlmeier, and Dr. Kathryn Braund for their work on my doctoral committee. Each of these professors has had a profound impact on my life in some way. This is especially true of Dr. Saye who has been my mentor since 1992. I've been truly blessed to learn from someone of his caliber and intellect. I owe a special debt of gratitude to Dr. Shannon for his assistance with the statistical portions of this research.

My family played an important role in helping me with this achievement. I would like to thank my parents (Jerry and Roberta Maddox), Ryan Maddox, Kathryn Maddox, and Jolie Maddox for their support. I'd also like to thank the Brenton family for allowing me the space and time to write during several holidays I spent with them in Indiana.

I would be amiss if I didn't also recognize my peers in the doctoral program at Auburn. I'd like to thank Dr. Charles Farmer, Dr. Linda Mitchell, Dr. Cory Callahan, Jay

Howell, Colby Jones, and Blake Busbin for their thoughtful advice and assistance at different stages of my graduate career. I'm especially grateful to Jay and Colby for setting aside time from their busy schedules to evaluate student essays and participate in lengthy inter-rater reliability sessions. I could not have completed this study without their support. My growth as a professional was also highly influenced through interactions with numerous graduate and undergraduate students I had the privilege of meeting while at Auburn University. I'd like to thank this group collectively for the positive impact they've had on my life.

Finally, I'd like to dedicate this work to my wife, Jennifer. Jennifer was daring enough to marry me while I was in the midst of this research. I'd like to thank her for tolerating my busy schedule and helping me to maintain my sanity during the most difficult stages of this study. Her support has been unwavering and it has made my success in this endeavor possible.

Style manual used is Publication Manual of the American Psychological Association, 6<sup>th</sup> Edition. Computer software used is Microsoft Word for Windows.

## Table of Contents

Abstract .....	ii
Acknowledgments .....	iii
List of Tables.....	ix
List of Figures .....	xi
List of Abbreviations .....	xii
CHAPTER ONE: INTRODUCTION .....	1
Study Overview and Methodology .....	5
Definitions.....	8
Study Limitations.....	10
Keywords .....	11
CHAPTER TWO: LITERATURE REVIEW .....	12
Theoretical Foundations.....	16
Learning Theory .....	16
Affordances of Disciplined Inquiry .....	19
Reservations .....	29
Authentic Intellectual Work .....	31
Research .....	34
Overview .....	34
Harvard Social Studies Project .....	42

Survey-Based Research.....	45
Learning Outcomes: Authentic Intellectual Work (AIW) .....	47
AIW and Lower Order Outcomes.....	48
AIW and Higher Order Outcomes .....	53
Gates Foundation Research.....	60
International Research.....	67
Adding to the Research base .....	73
CHAPTER THREE: METHODOLOGY .....	77
Study Design .....	79
Project setting and description of participants .....	80
Instrumentation.....	86
Study Phases.....	111
Data Analysis Procedures.....	119
Conclusion.....	130
CHAPTER FOUR: TEACHER USE OF AUTHENTIC PEDAGOGY .....	131
Minimal Authentic Pedagogy.....	136
Limited Authentic Pedagogy.....	146
Moderate Authentic Pedagogy .....	161
Generalizations .....	173
CHAPTER FIVE: STUDENT LEARNING OUTCOMES.....	180
Description of the Sample .....	180
Results of Inferential Analyses.....	182
Summary .....	207

CHAPTER SIX: SUMMARY, LIMITATIONS, & IMPLICATIONS.....	210
Summary .....	211
Discussion and Alternative Explanations.....	215
Limitations .....	221
Implications and Areas for Further Study .....	223
Conclusion.....	228
References.....	230
Appendix A: Teacher Interview Script.....	262
Appendix B: Teacher Recruitment Script.....	265
Appendix C: Scoring Criteria for Classroom Instruction .....	267
Appendix D: Scoring Tips for Instruction Rubric .....	269
Appendix E: Scoring Criteria for Tasks .....	271
Appendix F: Scoring Tips for Task Rubric.....	272
Appendix G: Email Correspondence Request for Tasks .....	273
Appendix H: U.S. History Higher Order Assessment Resources .....	274
Appendix I: U.S. History Higher Order Assessment Instructions .....	276
Appendix J: Advanced Placement Higher Order Assessment Student Resources.....	278
Appendix K: Advanced Placement Higher Order Assessment Instructions.....	281
Appendix L: Proctor Instructions .....	283
Appendix M: Scoring Rubric for Advanced Placement Higher Order Editorial .....	284
Appendix N: Scoring Rubric for Manifest Destiny Higher Order Assignment.....	290
Appendix O: Authentic Pedagogy Scores.....	296
Appendix P: Manifest Destiny Painting .....	299

Appendix Q: WWII Political Cartoon .....	301
Appendix R: Moderate Authentic Pedagogy Task .....	302
Appendix S: Content Analysis Explanation and Examples .....	304
Appendix T: Notes on the Student Sample .....	308
Appendix U: Technical Description of Multiple Regression Analysis .....	309
Appendix V: Technical Description of One-way ANOVA Procedures .....	311
Appendix W: Technical Description of Factorial MANOVA Procedures .....	313
Appendix X: Higher Order Editorial Examples .....	316



## List of Tables

Table 1: Summary of Results from Authentic Intellectual Work Studies .....	58
Table 2: Summary of AIW Studies with an Explicit Focus on Disadvantaged ..... Students	59
Table 3: Gates Foundation Studies: Authentic Student Work and Performance on ..... Standardized Tests	65
Table 4: Gates Foundation Studies: Relation between Authentic Tasks and Student..... Work	66
Table 5: Summary of International Studies .....	71
Table 6: Comparison of Tenth Grade Graduation Exam Passage Rates .....	82
Table 7: Descriptive Statistics for Teacher Sample .....	84
Table 8: Student Participation by Course .....	85
Table 9: Summary of Inter-Rater Reliability Observations .....	91
Table 10: Inter-Rater Agreement on Instruction and Assessment Tasks .....	91
Table 11: Inter-Rater Agreement on Higher-Order Editorial Tasks .....	103
Table 12: Summary of Inter-Rater Reliability Sessions .....	118
Table 13: Summary of Research Questions and Data Analysis Methodology .....	120
Table 14: Overview of Independent Variables Used During Regression Analyses.....	124
Table 15: Teacher Profiles .....	132
Table 16: Cut Scores.....	133
Table 17: Overview of Roy’s Authentic Pedagogy Scores .....	137
Table 18: Scores for “Teach a Lesson” Task.....	140

Table 19: Overview of Amy’s Authentic Pedagogy Scores .....	147
Table 20: Scores for “Ideal Form of Government” Task.....	150
Table 21: Phillip’s Authentic Pedagogy Scores.....	159
Table 22: Ryan’s Authentic Pedagogy Scores.....	164
Table 23: Scores for “WWII Political Cartoon Analysis” Task .....	166
Table 24: Scores for “Truman Think Aloud” Task .....	173
Table 25: Teacher Profiles .....	181
Table 26: Descriptive Statistics for Student Sample .....	181
Table 27: Impact of Authentic Pedagogy on Graduation Exam Results .....	183
Table 28: One-way ANOVA comparing Graduation Exam Scores for Minimal &.....	184
Limited Classes	
Table 29: One-way ANOVA Comparing Graduation Exam Scores for Minimal &.....	186
Moderate Classes	
Table 30: Distribution of Manifest Destiny Editorial Scores .....	189
Table 31: Comparison of Authentic Pedagogy Groups on Manifest Destiny Editorial .	190
Table 32: Distribution of German Unification Editorial Scores .....	194
Table 33: Comparison of AP Groups on German Unification Editorial .....	195
Table 34: Analysis of the Impact of Moderate Authentic Pedagogy on Graduation.....	200
Exam Results	
Table 35: Sequential Multiple Regression Analyses Predicting Impact of Repeated ..	203
Exposure to Moderate Authentic Pedagogy on Graduation Exam Results	
Table 36: Authentic Pedagogy and Achievement by Subgroups .....	205
Table 37: Authentic Tasks and Achievement by Subgroups .....	205
Table 38: Authentic Instruction and Achievement by Subgroups .....	206

## List of Figures

Figure 1. Process for determining Authentic Pedagogy Scores. ....	78
Figure 2. Summary of Research Phases.....	112
Figure 3. The “Teach a Lesson” Task .....	140
Figure 4. Reformers of the 1800s Task .....	145
Figure 5. Examples of Supporting Arguments for German Unification Editorial .....	198
Figure 6. Effect of repeated exposure to moderate authentic pedagogy.....	201

## List of Abbreviations

AHSGE	Alabama High School Graduation Exam
AIW	Authentic Intellectual Work
AP	Advanced Placement or Authentic Pedagogy depending on context
ARMT	Alabama Reading and Mathematics Test
CORS	Center on Restructuring Schools
IRB	Institutional Review Board
NAEP	National Assessment of Educational Progress
NCLB	No Child Left Behind
NCSS	National Council for the Social Studies
SAT-10	Stanford Achievement Test
SSIRC	Social Studies Inquiry Research Collaborative

## CHAPTER ONE: INTRODUCTION

What should students learn and how should they learn it? This question is a difficult one to answer regardless of the field, but especially when it comes to history. Efforts to devise standards in U.S. history have often resulted in heated debate and controversy whether at the state or national levels (Cheney, 1994; Symcox, 2002). The debate over what students should learn and how they should learn it in history is complex (Evans, 2004). Many people agree that the history curriculum, as part of the social studies, is vital for preparing good citizens. However, people have differing conceptions of America's democracy and the role of the "good citizen" within this context. As a result, a variety of curriculums have developed over time to educate secondary history students with different civic outcomes in mind. The following paragraphs provide a brief survey of three commonly known instructional approaches as the basis for discussing the impact of high-stakes testing on student learning.

Traditional instruction represents the oldest and most commonly used approach for teaching history. Students are asked to remember important names, dates, and events from the past as highlighted by the teacher or the textbook. Emphasis is often placed on student mastery of one main narrative of the past. This narrative tends to be a celebratory one depicting the steady progress of America's democracy (Barton & Levstik, 2004). The main goal of traditional instruction as it pertains to citizenship is often to instill

patriotism and cultural literacy. Instruction is mainly geared towards building foundational knowledge based on the belief that this is needed before significant higher order thinking can really take place (Hirsch, 1988; Newmann, Bryk, & Nagaoka, 2001, p. 11).

A second approach for teaching history has been especially popular since the 1960s. Advocates of disciplined inquiry believe students need to have the opportunity to “do history” using the techniques of historians in order to formulate more in-depth and nuanced understandings of the past (Seixas, 2001; Wineburg, 2001). In doing history, students might construct narratives of a particular historical event based on the analysis of primary sources. Engaging in historical interpretation has the potential to help students conceptualize the discipline in a manner that is more consistent with professional historians. Advocates of disciplined inquiry argue that this approach is more likely to help students develop the higher order thinking skills and dispositions needed for life in the 21<sup>st</sup> century. These learning outcomes are assumed to have civic value. For example, when students encounter a civic problem they should be able to apply their historical thinking skills to locate relevant information, evaluate its trustworthiness, analyze competing sources, and work through the problem to construct a supportable solution.

Finally, some social studies educators advocate problem-based historical inquiry (PBHI) directed towards the study of persistent issues affecting democracies (Saye & Brush, 2004) . Advocates of PBHI believe that in order for the knowledge, skills, and dispositions acquired in history classes to transfer to life outside of school, inquiry needs to be situated in real world social problems. For example, a unit on the Mexican-

American War might focus on whether the United States was justified in going to war with Mexico. In examining that historical problem, students would also consider the broader question of when one nation is justified in imposing its will on another. Criteria developed to address this broad question could be applied to the historical case of the Mexican-American War as well as other historical and contemporary conflicts. PBHI units are intended to help students see connections between the events they study in history and life in today's world. The focus on applying historical knowledge in realistic decision-making activities is designed to prepare students to be active citizens who can make decisions for the public good.

As can be seen through this brief overview, the social studies curriculum can be conceptualized in a variety of ways. Many social studies educators have been taught some version of inquiry-based instruction as a method to use with students to promote higher order thinking and other learning outcomes. It continues to be highly advocated through research publications and professional development initiatives. Despite attempts to influence the practice of teachers, inquiry-based instruction remains more highly regarded for its potential than its actual widespread use in schools. Inquiry-based instruction is difficult to implement and a variety of obstacles exist in school settings to limit its use (Rossi, 1998). This study focuses on one of the biggest disincentives to inquiry-based instruction – high stakes testing.

In many states social studies teachers must prepare their students to pass high-stakes standardized tests that primarily measure students' acquisition of lower order content knowledge. The tests often seem to be aligned to standards that reflect the goals of the traditional history model of instruction. They focus on how well students can

remember discrete facts from across the curriculum. The dilemma facing social studies teachers in this situation is an old one: depth vs. breadth. The high stakes tests seem to demand rapid coverage of information in order to ensure students are exposed to all of the testable content during a course. However, teachers who adopt more ambitious instructional goals are likely to favor in-depth treatment of specific historical topics in order to promote higher order learning outcomes. The concern among these teachers, of course, is whether their students will be able to pass the high stakes tests.

Advocates of inquiry-based instruction have argued that students learn just as much lower order content knowledge while engaged in active, inquiry-based activities as students in more traditional classroom settings. There is evidence to suggest this is true in other subjects, but the social studies research is not as strong. This study is an attempt to better ascertain some of the learning outcomes that can be expected from inquiry-based instruction in history classrooms. Hopefully the results of this study will offer some evidence to reassure teachers that inquiry-based instruction does no harm when it comes to student performance on the high-stakes tests that could determine their graduation status.

In order to conceptualize instruction in this study as a variable, I've used Newmann's authentic pedagogy framework. Teachers who use authentic pedagogy engage students in activities that require construction of knowledge, using elaborated forms of communication to create products that have value beyond school (Newmann, King, & Carmichael, 2007). This is their dominant practice. However, they also utilize more traditional instructional strategies such as lecture and multiple-choice tests as needed. In using Newmann's authentic intellectual rubrics to analyze instruction, I was



able to classify the teachers in this study on a continuum. Teachers on the lower end of the authentic intellectual work continuum use a great deal of didactic instruction. As the scores increase on the continuum, they represent greater use of authentic pedagogy (in-depth analysis of topics, inquiry, etc.).

Using this framework enabled me to overcome some of the problems that have historically plagued studies that have attempted to compare learning outcomes associated with traditional and inquiry-based instruction. Prior studies have compared inquiry classes with control classes. However, it was often hard for consumers of this research to determine the nature of intellectual challenge that was really present in the inquiry based classrooms. How different were they really from the traditional classes? In this study, all of the classes were assigned scores using the same task and instruction rubrics. This makes it easier to readily compare the degree of intellectual challenge experienced by students taught by one teacher as compared to another teacher in the study. It provides a better basis for comparing learning outcomes.

### **Study Overview and Methodology**

This was a mixed methods investigation of existing instruction at the study schools that involved collecting qualitative data and converting it to quantitative data for analysis. It also included the analysis of quantitative data in the form of test scores. I selected a junior high school and high school in southeastern Alabama as the focus schools for this study. I recruited the entire 9<sup>th</sup> and 10<sup>th</sup> grade social studies faculty as study participants. These teachers were asked to provide three challenging tasks that provided the best evidence of students performing their subject at the highest levels. I then established an observation schedule to coincide with the period when students would

be engaged in work related to the tasks. My analysis of the tasks and instruction, using Newmann's authentic intellectual work rubrics, resulted in each teacher being assigned an authentic pedagogy score. Cut scores were developed to form descriptive categories representing different levels of authentic pedagogy (minimal, limited, moderate, substantial). These data were used as the basis for an analysis of the impact of this type of instruction on student learning.

The student sample included four cohorts of tenth graders. I obtained achievement records, graduation exam results, and demographic data for these students during the 07/08 and 08/09 school years. All tenth grade students who took social studies courses during this time period were included in the study. The main instrument for measuring the retention of lower order social studies knowledge was the Alabama High School Graduation Exam (AHSGE). In order to measure higher order thinking, I created a writing assessment that measured the ability of students to analyze historic documents, formulate arguments, and make reasoned decisions. The higher order measure was administered to a smaller slice of the 10<sup>th</sup> grade student population.

I used several types of statistical analyses to determine the impact of authentic pedagogy on student learning outcomes on the AHSGE and the higher order essay. In doing so, I controlled for demographic and prior achievement variables likely to influence student performance. Once these variables were controlled for, the importance of instructional experiences in promoting the desired learning outcomes became more apparent.

**Research Questions.** The focus of this study was to examine the learning outcomes associated with various levels of authentic pedagogy. The research was guided by the following research questions:

Question 1: To what extent do teachers utilize authentic pedagogy and how much variation exists within the sample of teachers in this study?

Question 2: Do students that have been taught by teachers demonstrating higher levels of authentic pedagogy score higher on the Alabama High School Graduation Exam (AHSGE) than students taught by teachers with lower levels of authentic pedagogy?

Question 3: What is the impact of authentic pedagogy on student performance on an assessment that requires them to apply knowledge from a previous unit to a challenging new task?

Question 4: Does the ability to apply knowledge in these situations improve with repeated exposure (multiple courses) to classroom experiences that require students to perform challenging intellectual tasks?

Question 5: To what extent does authentic pedagogy bring different achievement benefits to students of different social and academic backgrounds?

**Study Purpose.** There is a lack of information within the research and literature describing the impact of authentic pedagogy on student learning in social studies. The purpose of this study is to better understand how the work students do in their social studies classes relates to their ability to apply what they learn on tests of lower and higher order knowledge. The study is timely and needed within the field. Today's high stakes

testing environment, which tends to focus on student acquisition of basic content knowledge, serves as a disincentive to teachers interested in using disciplined inquiry as part of authentic pedagogy. Teachers need to be able to turn to a body of research to support their use of authentic pedagogy under these circumstances. Information generated from this study could possibly contribute to the national body of research that suggests that problem-based historical inquiry not only helps students improve their critical thinking abilities, but also results in the knowledge needed to perform well on standardized tests.

### **Definitions.**

A number of terms are used throughout this study that may be new or ambiguous to some readers. In this section I have operationalized several of the most commonly used terms.

**Authentic Pedagogy.** Authentic pedagogy includes any instructional practices designed to elicit authentic intellectual work from students. A teacher's pedagogy, according to Fred Newmann, is a combination of daily instruction and assessment tasks. In order for a teacher's pedagogy to be considered "authentic" it must adhere to certain standards. Authentic instruction is designed to promote higher order thinking, depth of knowledge, substantive communication, and a connection to life outside the classroom. Authentic tasks are designed to promote construction of knowledge, elaborated communication, and a connection to students' lives.

**Authentic Intellectual Work.** Fred Newmann juxtaposes the work students are traditionally asked to complete in school with work he considers to be "authentic". Whereas traditional school assignments are often used to simply certify success in school,

authentic achievements have broader personal significance and meaning in the real world. As such, they closely mimic the thinking and effort required of significant intellectual accomplishments for adults. Students engage in authentic intellectual work when they “construct knowledge, through disciplined inquiry, to produce discourse, products, or a performance that has value beyond school” (Newmann, King, & Carmichael, 2007, p. 3).

**Disciplined Inquiry.** In this study, inquiry is considered disciplined when it adheres to the conventions and methods of reasoning associated with a particular field of study. In other words, when students engage in problem-solving in history they must produce defensible solutions that would be seen as valid among professional historians. Disciplined inquiry requires students to develop a knowledge base, strive for in-depth understanding, and communicate ideas using elaborated forms of communication (Newmann, King, & Carmichael, 2007).

**Traditional Instruction.** The primary purpose of traditional instruction is the delivery of content which students are asked to remember and recite (Lee, Smith, & Newmann, 2001, p. 10). Traditional instruction is teacher-centered and dominated by lecture and drill and practice exercises. This term is most often used in this study to describe instructional practices that generally do not follow the standards associated with the authentic pedagogy framework.

**Standardized test-based reform.** My conception of standardized test-based reform stems from Scott Thompson’s use of the term. This is a system of reform where “academic progress is judged by a single indicator and when high stakes – such as whether a student is promoted from one grade to the next or is eligible for a diploma – are attached to that single indicator...(Thompson, 2001, p. 358).”

### **Study Limitations.**

This study has several potential limitations. The first limitation relates to the format of the Alabama High School Graduation Exam. The exam covers U.S. history from exploration through World War II. Students in the focus grade of this study (grade 10), regardless of their instructional experiences, did not have the opportunity to learn all of the testable content. The tenth grade course only covers the first half of the U.S. history survey. The graduation exam is provided to tenth grade students mainly as a familiarization exercise, although those who pass do not have to take the test again. Ideally, the lower order content knowledge measure for this study would have more closely adhered to the curriculum students experienced. As such, the passage rates on this test could have been influenced, to a greater extent than usual, by non-instruction related factors such as the education level of a student's parents.

Another graduation exam related limitation stems from the difficulty I encountered in conducting a content analysis of the test. A content analysis was needed to verify that the test items were predominately focused on measuring lower order content knowledge. The state of Alabama does not provide public access to tests or test questions used in previous years. A bulletin with 84 sample test items was the only thing available from the state. I used this bulletin for my analysis despite the fact that there was no assurance that the sample items were comparable to those found on the actual graduation exam. This made it difficult to determine the challenge level of the test with absolute certainty.

Another study limitation relates to the collection of data. Ideally, I would use all of Newmann's authentic intellectual work rubrics (task, instruction, and student work) to determine levels of authentic pedagogy provided by the sample of teachers. However, I simply did not have the resources available to collect and analyze the student work associated with the tasks assigned by the study teachers. An analysis of student work would have been useful to gain a better sense of the degree to which students were engaged in such standards as construction of knowledge and elaborated communication.

Finally, a potential limitation involves the ability to make generalizations from this study. This study includes a very limited sample of teachers and uses outcome measures not found in other states. However, it is a pilot study for a larger effort by the Social Studies Inquiry Research Collaborative (SSIRC) focused on essentially the same research questions. This association with the work of other researchers will hopefully allow the results to be more meaningful.

### **Keywords**

Authentic Intellectual Work, Assessment, Social Studies Education Reform, Inquiry-based instruction

## CHAPTER TWO: LITERATURE REVIEW

Many people believe public schools are not doing an adequate job of preparing students for life in the 21<sup>st</sup> century (Partnership for 21st Century Skills, 2007). In order to remedy this situation, a variety of reform initiatives have been suggested. Each of these is designed to influence the quality of instruction in some way. The *No Child Left Behind* (NCLB) test-based accountability model uses rewards or sanctions based on standardized test results to improve instruction. A similar reform initiative uses value-added statistical modeling to hold teachers accountable for how much students learn during a semester (Braun, 2005; Koedel & Betts, 2009; Rothstein, 2009; Stewart, 2006). A third reform model is based on the authentic pedagogy construct devised by Newmann (Newmann & Archbald, 1988; Newmann, King, & Carmichael, 2007; Newmann, Secada, & Wehlage, 1995). Supporters of authentic pedagogy seek to improve the capacity of teachers to provide intellectually challenging instruction according to standards of Authentic Intellectual Work (AIW). A wide range of other ideas for improvement have been proposed to include additional coursework requirements and calls for more active instruction (Smith & Niemi, 2001).

The mainstream reform model today is *No Child Left Behind* (No Child Left Behind Act of 2001, 2002). While NCLB focuses primarily on improving math and reading achievement, many states have adopted test-based reform as an accountability



measure for social studies. Advocates of this system believe high-stakes attached to tests will improve student motivation, effort, and achievement (Stecher, 2002). The tests are also meant to apply pressure on schools and teachers in a variety of ways. It is believed that if students fail, teachers will work harder to improve their instruction so the standards are accomplished. The standards and tests adopted by states send a message to teachers regarding the types of learning outcomes that are most valued. However, reports by a variety of think-tanks and policy organizations consistently criticize many of the graduation exams and high stakes tests for their lack of rigor (Achieve Inc., 2004; Conley, 2003; Cronin, Dahlin, Adkins, & Kingsbury, 2007; Daugherty, 2004).

Many states use high-stakes history tests that emphasize lower order outcomes (Gaudelli, 2006; Grant & Horn, 2006). These multiple-choice tests are often designed to see if students know “the basics” (Newmann, Bryk, & Nagaoka, 2001). Tests of basic knowledge usually assess the ability of students to remember specific factual information (names, dates, events). In this type of environment, teachers feel pressured to adopt coverage-based instructional approaches to survey all the possible content students might encounter (Grant, 2005; Grant et al., 2002). If teachers are going to pursue the type of in-depth, inquiry-based instruction advocated by many researchers, they need evidence that it will not hurt their students on these tests (Grant et al., 2002).

The dilemma facing teachers in the present high-stakes environment highlights a longstanding controversy in the social studies field. Should social studies courses be survey-oriented or should they provide students with in-depth learning experiences

(Newmann, Lopez, & Bryk, 1998; Parker, 1991; Rossi, 1995; Rothstein, 2004)? These two instructional approaches are based on very different assumptions regarding the purpose of social studies and what constitutes meaningful instruction. Traditional survey instruction is usually focused on transmitting factual knowledge to students while in-depth instruction is more concerned with obtaining higher order thinking objectives (Rossi, 1995). It therefore seems likely that the goals of a broad, coverage-oriented survey course would more closely align with the format of many high-stakes history tests. However, proponents of in-depth, inquiry-oriented instruction argue that this type of instruction is more effective in helping students achieve both lower and higher order outcomes. This chapter presents the theoretical argument for and against this statement as well as empirical research that has tested this claim.

I begin with a basic explanation of in-depth instruction. Rossi's operational definition of in-depth instruction provides a clearer picture of how this term has been conceptualized in social studies. The construct encompasses issues-centered, inquiry-based instruction and involves:

1. The use of knowledge that is complex, thick, and divergent about a single topic, concept, or event using sources that range beyond the textbook;
2. Essential and authentic issues or questions containing ambiguity, doubt, or controversy;
3. A spirit of inquiry that provides opportunities, support, and assessment mechanisms for students to manipulate ideas in ways that transform their meaning; and
4. Sustained time on a single topic, concept, or event. (Rossi, 1995, p. 89)

In-depth units can take a variety of forms under this broad definition. Implicit is the idea that instruction should foster understanding and the ability of students to think.

Most social educators believe in-depth instruction is necessary if students are to become effective citizens who are able to apply what they've learned to make decisions for the public good in a diverse society (National Council for the Social Studies, 1994).

In-depth units represent a departure from traditional coverage-based approaches to instruction in a number of ways. In traditional instruction, the teacher is primarily concerned with the student producing the right answer. The teacher's role is to transmit factual information to the student. The student is then tested on his/her ability to reproduce this information (Lee, Smith, & Newmann, 2001). In-depth inquiry units, on the other hand, typically require construction of knowledge. The end goal of in-depth instruction, as described by Rossi and others, is not to determine how many facts students can remember (although facts are still considered important). It is to evaluate the quality of students' reasoning and understanding according to the intellectual standards of a particular discipline. How well can students marshal relevant facts to support an argument?

Having established the definition of in-depth instruction and its relation to traditional instruction, I now present some of the major reasons for its use in schools. Advocates support in-depth instruction because it is grounded in disciplinary standards, can be used to promote the citizenship mission of social studies, and is consistent with contemporary understandings of how people learn. Each of these points will be examined more closely in the following sections. Because this study focuses only on instruction in history classrooms, the terms in-depth instruction and historical inquiry are used interchangeably. However, Rossi's definition applies across the social science disciplines.

## **Theoretical Foundations**

### **Learning Theory**

It seems counter-intuitive that in-depth units would result in the type of learning necessary for students to excel on tests of basic factual knowledge. After all, an in-depth curriculum usually involves sacrificing some breadth. How can students pass tests that cover a broad range of content? Advocates believe constructivist theories of learning help to explain the effectiveness of in-depth instruction.

Constructivists view students as active meaning-makers in the instructional process who interpret what happens in a classroom environment based on their prior knowledge and experiences (Bransford, 2000; Brooks & Brooks, 1993; Scheurman, 1998). They internalize information into mental maps or “schemas” (Bartlett, 1932; Piaget, 1952; Rumelhart, 1980). When new information is presented during class, students either add it to an existing schema, modify a schema to accommodate it, or create an entirely new schema if the information differs radically from what they have experienced previously (Cornbleth, 1985; Rumelhart, 1980).

The development of complex schemata and deep knowledge is thought to be the key to long-term memory as well as higher-order problem-solving (Greeno, Collins, & Resnick, 1996). Much of the basis for this belief comes from studies that analyze the different ways experts and novices in particular fields solve novel problems. These studies indicate that expert knowledge is organized hierarchically around big ideas and concepts while the knowledge of novices tends to be fragmented. Knowledge that is highly connected in this manner is more accessible for rapid recall and can be flexibly

applied for problem-solving (Bransford, 2000; VanSickle & Hoge, 1991; Wineburg, 1991). In contrast, much of the knowledge of novices is inert (Whitehead, 1929). This simply means that novices are often unable to recognize when their knowledge might be applicable when confronted with problems they haven't experienced previously (Cognition and Technology Group at Vanderbilt, 1990).

The research on expert/novice problem-solving supports the need to have deep, complex knowledge. The ability of students to form these complex connections is thought to depend on the nature of the instructional experience they encounter. The question therefore becomes: Which instructional approach is more likely to promote the understanding necessary for the development of complex schemata? Advocates of traditional survey-oriented instruction tend to place the greatest emphasis on drill and repetition (Greeno, Collins, & Resnick, 1996). Students are taught factual information and then subsequent courses add more information to the knowledge base. The intent is for instruction to have a cumulative effect over time to where students eventually form more complex understandings. Constructivists argue that this approach is based on a faulty belief of how humans learn. If knowledge is never successfully integrated into students' schemata, it is likely to result in disconnected rather than connected knowledge (Bransford, 2000, p. 30).

Advocates of in-depth instruction believe it promotes the development of complex schemata in a number of ways. A good central question or problem captures students' attention and promotes a "felt need" to resolve an issue (Dewey, 1938). Authentic questions, those that have relevance in contemporary life, provide students with a greater purpose for engaging in the learning process. The issue or problem in a lesson becomes

the focal point around which students organize information; the mental peg which aids in memory and application of knowledge (Brooks & Brooks, 1993). As students investigate a problem, they are forced to actively work with information and reorganize it in new ways. The resulting process of knowledge construction (in solving problems) is believed to help students gain depth of knowledge and the ability to think at a higher level.

In-depth instruction provides skilled teachers with a greater opportunity to diagnose student misunderstandings and provide support for learning. Vygotsky believed students had a “zone of proximal development (ZPD)” which constituted the difference between what students could do on their own and what students could do with guidance (Vygotsky, 1978). Based on this idea, educators advocated the use of scaffolds to enhance student learning. The use of scaffolds in the classroom can be compared to spotters in weight-lifting. Spotters help athletes lift heavier weight and complete more repetitions than they ever could on their own. The good ones make the athlete do the bulk of the work providing just enough support to complete the exercise. Eventually, the lifter progresses to where he/she can lift the weight without the support. The same idea applies to scaffolding in the classroom. The teacher locates the student’s current developmental level and seeks to provide support (questions, sequenced activities, etc.) to help students stretch their intellectual abilities. In-depth instruction allows teachers to intervene with scaffolding to optimize learning and help students develop rigorous (discipline-based) solutions to problems (Scheurman, 1998; Vygotsky, 1978).

In summary, many social educators endorse in-depth, inquiry-based instruction because they believe it allows students to build more complex understandings of historical concepts than is possible in coverage based environments. This belief is

supported by the research of many cognitive scientists. Deep knowledge and understanding is thought to be the key not only to improved problem-solving capacity, but also efficient recall of factual information. If standardized tests are mainly tests of reading comprehension as some suggest (Newmann, Bryk, & Nagaoka, 2001), students that have developed complex schemata should possess a network of associations that can be used to more effectively make inferences when reading and answering multiple choice questions (Doyle, 1983, p. 167; Nuthall & Alton-Lee, 1995).

### **Affordances of Disciplined Inquiry**

The close relationship between in-depth instruction and contemporary understandings of how people learn provides one justification for the use of inquiry-based instructional practices in social studies. Many researchers, educators, and policy makers also support inquiry because it is central to the practice of social scientists and therefore essential for helping students developed more accurate understandings of the discipline under study (Barton & Levstik, 2004; Wineburg, 2001, Seixas, 2001). Since my research dealt with instruction in history classrooms, examples in this section will focus on this field and how inquiry is applied within research on issues-centered curriculums.

Wineburg's research has been very influential among educators seeking to help students learn to think historically. One of his most well known studies compared high school students and professional historians as they reasoned with historical texts (Wineburg, 1991). This study revealed major differences in the way historians and students viewed historical knowledge. The students viewed historical accounts as definitive and truthful and this limited their ability to recognize the need to look for underlying meanings and subtext in the documents they were provided during the study.

The historians, meanwhile, viewed the same accounts as “human creations” requiring analysis and interpretation to fully understand (Wineburg, 1991, p. 510-12). This led them to apply sourcing, contextualization, and corroboration skills to develop reasoned conclusions about the trustworthiness of the documents. Wineburg believed that the students’ inability to reason deeply with the texts was primarily due to the textbook driven history they likely encountered in school (see Baldi, et al., 2001).

Textbooks often portray history as a single meta-narrative. This creates the illusion that historical knowledge is static and relatively uncomplicated (Bain, 2000; Gabella, 1994; Wineburg, 1991). Students gain little sense of the scholarly debate surrounding many of the topics they study. Inquiry-based instruction has the potential to shift the epistemological stance of students from the perception of history as something to be memorized, to an “uninterrupted negotiation about the character of the past” (Nash, 1995, p. A2).

In addition to developing a more discipline-based understanding of history, inquiry advocates believe this instructional approach provides students with a broad range of historical thinking and reasoning skills that have application in the real world. When students engage in historical inquiry they get to “do” history. They analyze sources of evidence to develop their own account of an historical event. There are some inherent dangers in this process. Without a proper understanding of the rules of evidence used by historians, students could form unwarranted conclusions or develop shallow interpretations of the past. Teachers also have to fight the tendency of students to become relativistic once they understand that the past can be viewed from multiple perspectives (Saye, 1999; Barton, 2008). The upside of allowing students to engage in



inquiry is that with proper guidance, students learn firsthand how historical knowledge is constructed. They also develop skills such as the ability to critically examine sources of evidence, detect bias, make logical inferences and generalizations, evaluate the trustworthiness of competing accounts, synthesize information, look at problems from multiple perspectives, and empathize with the perspectives of people from different times, places, and cultures (Barton, 2008; Kohlmeier, 2006; Saye & Brush, 2007; VanSledright, 2002).

Students in traditional classroom settings are also exposed to many of these skills, but this is typically through worksheets or more limited classroom exercises that are sometimes isolated from the primary objectives of an instructional unit. Inquiry advocates believe historical thinking skills are difficult to learn out of context. In inquiry-based classes, teachers embed these skills within major instructional activities. When students engage in realistic inquiry activities, advocates believe they are more likely to then be able to apply these skills in the real world (Brown, Collins, & Duguid, 1989; Cognition and Technology Group at Vanderbilt, 1990).

Studies suggest that students can be taught historical thinking skills and the ability to formulate reasoned decisions about contemporary social issues, even at a relatively early age (Barton, 1997; Foster & Yeager, 1999; Lee & Ashby, 2000; Saye & Brush, 2007; VanSledright, 2002). Researchers have documented improvements in these areas in a number of ways. Some observe classroom teachers to evaluate the effectiveness of various instructional practices while others evaluate the impact of specific interventions. The following paragraphs briefly describe the research basis for some of the more common claims made by inquiry advocates. The studies are organized into two

categories: historical thinking & reasoning and decision-making. This division reflects the primary orientation of the studies (i.e. historical inquiry vs. issues-centered instruction). In both categories, students engage in activities that can build skills and dispositions needed for effective citizenship. However, the citizenship focus is generally more explicit when looking at the issues-centered studies. The main purpose here is to highlight some of the major research outcomes that are often cited as affordances of having students engage in inquiry in their social studies classes.

**Historical Thinking & Reasoning.** Research by Young & Leinhardt (1998), Monte Sano (2008), De La Paz (2005), Ferreti, et al., (2002), and Kohlmeier (2005/06) support the idea that inquiry-based instruction can develop students' capacity to think and reason on tasks that require constructing evidence-based arguments. Young & Leinhardt (1998) & Monte Sano (2008) found a positive relationship to exist between classroom environments that emphasized inquiry and historical interpretation and the ability of students to construct evidence-based essays. Young & Leinhardt examined the effect of the document based questions (DBQs) commonly experienced by students in Advanced Placement courses on historical thinking. In their study, students became more adept at applying historical thinking skills on successive DBQs despite receiving little direct instruction on how to complete the task itself (Young & Leinhardt, 1998). The researchers attributed this improvement to the teacher's use of classroom activities that required students to construct arguments using a variety of different types of evidence. Monte Sano compared the instructional approach of two teachers and, in particular, their methods for teaching historical writing. Students in the class oriented around a more

inquiry-based instructional model wrote essays that demonstrated significantly greater levels of historical argumentation and reasoning. Other studies have noted increases in the ability of students to engage in historical inquiry in classrooms with similar characteristics (Gabella, 1994; Grant, 2001a).

Research by De La Paz, Ferreti, et al., and Kohlmeier featured more explicit interventions designed to elicit advanced historical thinking outcomes. De La Paz (2005) analyzed the ability of a diverse group of 8<sup>th</sup> grade students to apply historical inquiry skills after taking part in an integrated language arts/social studies unit. Students were broken into three groups for analysis: students with learning disabilities, average writers, and talented writers. After a relatively brief intervention of approximately two weeks, students in the experimental group constructed a document based persuasive essay. The essays were evaluated in terms of their length, persuasiveness, number of arguments, and accuracy. Students in the experimental group achieved higher scores in each of these areas than students in the control group. Students from each of the groups also showed improvement in the areas being measured when compared to their pre-test essays. Ferreti, MacArthur, & Okolo (2001) found similar results in a study that included eighty-seven fifth grade students in an urban, inclusion classroom environment. Participants in this study experienced an eight week project-based inquiry curriculum that concluded with students developing multi-media presentations describing the perspectives of particular groups involved in westward expansion. Students achieved statistically significant improvements over their pre-test scores in the areas of content knowledge and application of historical inquiry skills.

Kohlmeier (2005, 2006) used a three step instructional approach to help 9<sup>th</sup> grade World History students effectively reason with documents to develop a deeper understanding of the experiences of historical women. Students received sets of primary documents, at different points in the semester, describing the perspectives of women living during the Renaissance, Russian Revolution, and Cultural Revolution in China. Each time they encountered the source documents, they completed a reading web, Socratic seminar, and essay task. As the students gained experience interpreting history and constructing evidence-based essays over the course of a semester, they demonstrated a better understanding of the role of historians in creating historical narratives (Kohlmeier, 2005). Kohlmeier found that the documents and the three step process were successful in getting students to empathize with the perspectives of women and “ordinary people” during the periods under study. The three step instructional model improved the ability of students to critically analyze documents and write evidence-based essays (Kohlmeier, 2006). Perhaps most significantly, at least one of Kohlmeier’s students mentioned applying skills from the course to his own reading of contemporary articles that were not assigned in class (Kohlmeier, 2005).

These studies, taken in total, suggest that students can be taught to apply historical inquiry skills to document sets to construct reasoned arguments. In addition, Kohlmeier’s research shows the power of using carefully chosen documents in class to motivate students and cultivate historical empathy. The studies were not without their faults. In most cases they featured brief interventions and small sample sizes. In De La Paz’s study (2005), students who did not master important aspects of the experimental curriculum were excluded from the final analysis. Barton (2008) notes many of these same

weakness in his review of research on students' historical thinking, but argues that the consistency of findings among a diverse body of research is encouraging.

**Decision-Making.** Issues centered curriculums seek to foster some of the same types of thinking skills described in the previous studies, but they connect disciplined-inquiry to broader citizenship outcomes. Researchers are interested in whether students can apply these skills within the context of formulating reasoned decisions about contemporary societal issues (Engle, 1960). Since decision-making in a democracy doesn't occur in a vacuum, social inquiry in the classroom typically involves activities which require discussion and collaboration. When evaluating the effectiveness of interventions, researchers tend to use tasks (written & oral) that require persuasive argumentation (Newmann, 1990, 1991a; Parker, Mueller, & Wendling, 1989; Saye & Brush, 1999a). One such study by Newmann engaged students in a persuasive essay task based on a question involving the justification of a locker search. His research suggested that simple exposure to a classroom environment that exhibited general characteristics thought to promote higher order thinking was not enough to improve students' abilities to reason about an unfamiliar topic (Newmann, 1991c). However, later studies suggested that students can experience success on similar tasks with explicit scaffolding and support.

Parker, Mueller, & Wendling's (1989) study demonstrated the ability of high school students to engage in dialectical reasoning when asked to write an essay on a civic issue. The use of a scaffolded essay design encouraged the majority of the students to be able to argue both sides of the issue and empathize with opposing views (Parker, Mueller, & Wendling, 1989). Research by Saye and Brush built on these findings while

investigating the potential of technology to help students more effectively reason about social issues (Saye & Brush, 1999a, 2002). These researchers conducted a design experiment where they studied the effects of successive implementations of a problem-based unit on the Civil Rights movement. Even though the problem-based unit was designed and executed by a teacher with little experience with inquiry-based instruction, the students in the experimental class wrote essays that were more persuasive and featured higher dialectical reasoning scores than their peers (Saye & Brush, 1999a). When additional scaffolding was introduced with a new class in the second iteration of the study, students performed better than students from the 1<sup>st</sup> iteration in their ability to construct persuasive multi-media presentations that effectively used evidence to argue a position (Saye & Brush, 2002).

A wide range of additional civic outcomes have been documented by researchers who study the effects of controversial issues discussions (Hahn & Tocci, 1990; Hess & Posselt, 2002; Kahne & Spote, 2008; Larson, 2003; McDevitt & Kioussis, 2006; Torney-Purta, 2002). Hess & Posselt (2002) investigated how 10<sup>th</sup> grade students experienced controversial public issues (CPI) discussions over the course of a semester. Two teachers were observed as they implemented a curriculum that featured discussions related to five public issues. Students learned a variety of discussion skills such as how to ask probing questions, cite evidence to support an argument, make stipulations, and identify and explain value conflicts reflected in an issue. Through the analysis of a variety of data sources (i.e. interviews, scored discussions, class observations, questionnaires), Hess & Posselt concluded that CPI discussions improved the discussion skills of participants and that students generally liked engaging in this type of activity.

Another study conducted by McDevitt & Kiouisis (2006) also found positive outcomes associated with controversial issue discussions. This study evaluated longitudinal outcomes associated with the Kids Voting USA curriculum; a curriculum that includes service learning, mock election voting, family outreach activities, and other activities designed to inculcate deliberative habits in students. The researchers found the use of frequent classroom discussions, about election issues where students could express their opinions, to be among the most effective strategies for promoting long term civic development (McDevitt & Kiouisis, 2006). Survey and focus group discussions revealed several benefits of discussion to include “increased news attention, political conversations with parents, opinion formation, and motivation for voting” (McDevitt & Kiouisis, 2006, p. 4). The effects of the curriculum were shown to persist for two years after it was initially introduced resulting in “self perpetuating” habits associated with deliberative democracy. Other researchers have also noted that the discussion of controversial issues in an open classroom environment can promote civic engagement and participatory attitudes (Torney-Purta, J., 2002; Kahne & Spote, 2008; Hahn & Tocci, 1990).

Some researchers have analyzed inquiry curriculums designed to produce specific dispositions such as tolerance. Avery and her colleagues conducted a study with 274 9<sup>th</sup> grade students which evaluated an inquiry-oriented program designed to help students recognize the civil liberties of groups with whom they disagree (Avery, Bird, Johnstone, Sullivan, & Thalhammer, 1992). Analysis of the effects of this curriculum indicated that the students in the experimental groups experienced statistically significant increases in tolerance above and beyond those in the control. Avery concluded that a “curriculum

that helps students comprehend the consequences of intolerance can increase students' willingness to extend rights to disliked groups" (Avery, et al., p. 410).

**Summary.** Social educators tend to view classrooms as "laboratories of democracy" where students work together to make sense of societal problems (Parker, 1996). The inquiry process involves a number of steps which are not necessarily linear. It begins with the selection of a meaningful question. Historical thinking and reasoning skills are used to gather relevant foundational knowledge and evidence. Student understanding is further enhanced by classroom deliberations that reveal different views and perspectives on the problem. The outcome of these activities is the construction of an individual or group decision about the question. The studies reviewed in this section suggest a wide variety of beneficial outcomes can result from the inquiry process to include increased tolerance, participation in the political process, attention to news events, and an enhanced ability to develop reasoned positions on important issues.

A broad range of studies and interests fit under the "inquiry" umbrella in social studies. Despite persistent appeals by inquiry advocates, significant evidence suggests that students rarely experience this type of instruction (Baldi, Perie, Skidmore, Greenberg, & Hahn, 2001; Goodlad, 1984; Kahne, Rodriguez, Smith, & Thiede, 2000; Levstik, 2008; Rogers & Freiburg, 1994; Sizer, 1984). The lack of an overall consensus regarding inquiry or its purposes in social studies certainly makes it difficult for practitioners to envision alternative teaching strategies. Inquiry is also very challenging and time consuming making it less practical given institutional barriers that commonly exist in schools (Onosko, 1991). The next section explores some of the main reasons why some researchers argue for more traditional approaches to teaching social studies.



## **Reservations**

Opponents of inquiry-oriented instruction criticize the research described in the previous section for many reasons. Some researchers argue that controversial public issues instruction requires the ability to reason at levels that exceed most students' capabilities (King & Kitchener, 1994; Leming, 2003). Others have voiced concerns about the ability of teachers and students to reason effectively about the past without resorting to "presentistic" interpretations (Stern, Chesson, Klee, & Spoehr, 2003, p. 10). Finally, there is a general belief that students don't have the content knowledge base needed to engage in critical thinking (Onosko, 1991; Ravitch & Finn, 1987). These arguments are applied even more strenuously when discussing the capabilities of disadvantaged students and those with disabilities (Rossi, 1998).

Critics also question the wisdom of constructivism and the student-centered approach commonly associated with inquiry-based curriculums (Frazee & Ayers, 2003; Schug, 2003). Frazee & Ayers (2003) have argued that essential content gets shortchanged when teachers attempt to apply constructivist practices in their classroom. The consequences of shortchanging content, according to Hirsch, are most severely felt by disadvantaged students who miss out because they don't have the same learning opportunities outside of school as their more affluent classmates (Hirsch, 2009-2010).

Constructivist oriented curriculums face an uphill battle in winning over the general public. Traditional beliefs about learning remain entrenched in the public psyche (Powell, 1985, p. 311). The traditional learning paradigm holds that most students don't find academic work to be very interesting or motivating and therefore they must be pushed to achieve – especially through external reward systems (Brooks & Brooks,

1993). Attempts to revise the curriculum to tap into the intrinsic desire of humans to learn are often viewed with skepticism. People question whether these reforms are as rigorous as the instruction they might have received while in school (Newmann, Marks, & Gamoran, 1996; Windschitl, 2002).

Traditional notions about teaching are bolstered by researchers who claim that achievement can best be improved through direct instruction (Kirschner, Sweller, & Clark, 2006; Schug, 2003). Schug has argued that the poor performance of students on content knowledge tests like the U.S. history portion of the NAEP is a reflection of inadequate preparation of teachers. In his view, teachers jettison student-centered, constructivist pedagogy when they encounter the real world of the classroom, only to find they have no preparation in how to use direct instructional approaches that actually work (Schug, 2003, pp. 124, 127). The constructivist emphasis of teacher education is the culprit, rather than teacher-centered instruction which he acknowledges as dominant in most schools.

Perhaps the most fundamental difference of opinion, when it comes to history instruction, centers on the basic purpose of including this subject in the curriculum. Many advocates of history in public schools want students to learn traditional interpretations of the past (Cheney, 1994; Newmann, 1991a, p. 391). They argue that students need increased knowledge of U.S. history for the purpose of cultural literacy (Bennett, 1992; Hess, 2008b; Hirsch, 1988) and national unity (Finn, 2003; Paul Gagnon and the Bradley Commission on History in Schools, 1989; Saxe, 2003). Critical pedagogy is generally opposed because it is believed to undermine the more patriotic narrative of national progress found in many textbooks. The debate over the national

history standards and a Florida bill which defines American history for classroom instruction as factual and not constructed demonstrate the discomfort many Americans feel with postmodernism and curriculums designed to encourage historical analysis and interpretation (Laws of Florida, 2006; Symcox, 2002).

The merits of inquiry-based instruction have therefore been hotly debated for many years. Reconciling these competing perspectives often seems like an intractable problem. The next section describes the authentic intellectual work model - a vision for a more rigorous form of inquiry-based instruction that has gained the attention of many education reformers.

### **Authentic Intellectual Work**

The authentic pedagogy model proposed by Fred Newmann provides a framework that addresses at least some of the criticisms voiced by skeptics of inquiry-based instruction. Like NCLB, it is a product of the standards and accountability movement of the 1980s. However, rather than using high-stakes tests as a “lever” for instructional reform (Grant, 2001b), this model is designed to improve student learning outcomes by focusing on the quality of instruction. The major problem with in-depth programs, as described by Newmann and his associates, is the implementation of constructivist teaching strategies without standards of quality (Newmann, Marks, & Gamoran, 1996, p. 280). Teachers implement constructivist strategies (i.e. projects, hands-on activities, etc.) without ensuring the work students are asked to complete is rigorous and grounded in disciplinary standards (Newmann, Marks, & Gamoran, 1996). The authentic pedagogy model provides a framework that helps educators engage students in the types of intellectual work they are likely to encounter in today’s society. In this model,

“authentic” refers to school tasks that are complex enough to be considered “socially or personally meaningful”; on par with the types of intellectual accomplishments performed by adults (Newmann, King, & Carmichael, 2007, p. 2-3).

In developing this model, Newmann and his colleagues examined a wide variety of intellectual challenges encountered by people in their daily occupations to “define criteria for intellectual performance necessary for success in contemporary society (Newmann, King, Carmichael, 2007, p. 2).” They found that adults routinely face problems that require them to construct or develop solutions by applying what they know. Newmann concluded from this analysis that meaningful classroom instruction must move beyond memorization of factual material to provide students with similar intellectual experiences. While students are not expected to be on the same level as adults, Newmann’s vision is a curriculum where students are engaged in complex intellectual challenges that have importance beyond certifying success in school (Newmann, King, & Carmichael, 2007, p.5).

The theoretical basis and major components of authentic intellectual work (AIW) are discussed in several important works (Newmann, 1991a; Newmann & Archbald, 1988; Newmann, Secada, & Wehlage, 1995; Nystrand & Gamoran, 1990; Resnick, 1987; Wiggins, 1989). AIW includes the following main components: construction of knowledge, disciplined-inquiry, and value beyond school. Each of these components is described in further detail in specific standards.

The first component, construction of knowledge, requires students to move from being consumers of information to producers. They must use their prior knowledge and

information they learn in class to construct new (for them) interpretations or solutions to problems. This clearly involves significant higher order thinking.

The process used in developing solutions to problems is called disciplined-inquiry. Disciplined-inquiry is advocated to ensure students develop rigorous interpretations or solutions. This means that students must use procedures and “rules of evidence” that are considered legitimate by professionals in the academic discipline under study (i.e. historians, economists, etc.). A disciplined approach to inquiry also requires students to convey their findings to others through elaborated forms of communication. This can include a variety of formats to include more traditional essays or projects. The goal is for students to provide deep and nuanced explanations of their work.

Finally, authentic intellectual work has value beyond school. Student work that has value beyond school is focused on a real world problem and is often designed to “have an impact on others” (Newmann, King, & Carmichael, 2007, p.5). In social studies, an example might be if students tried to influence public policy by writing a persuasive letter to a Congressman (and actually sent it) or if a class created an informative website describing opposing perspectives on the issues for an upcoming election. These types of activities are meaningful and significant because students are grappling with the same types of intellectual challenges as adults. Ideally, the authenticity of the task evokes an emotional and personal investment in students as they strive to meet or exceed real world standards. Students know their work will be evaluated (informally or formally) by a public audience that is familiar with standards of excellence associated with the task. As band performances and athletic competitions

demonstrate, public scrutiny of this nature can motivate students to excel (Wiggins, 1993b). Newmann believes that teachers who offer students opportunities to construct knowledge, engage in disciplined-inquiry, and develop products that have value beyond school will have greater success in helping students obtain lower and higher-order learning outcomes that are authentic according to his definition.

This reform model is attractive to advocates of 21<sup>st</sup> century skills and other education stakeholders anxious to see high school students obtain the type of education that will allow them to compete in global, information-based economy (Kozma, 2008; Pink, 2008; Wallis & Steptoe, 2006). It also emphasizes disciplinary knowledge making it more palatable to proponents of the “basics.” The use of authentic pedagogy therefore offers a potential way to bridge the gap between proponents of instruction for higher order outcomes (the disciplined inquiry advocates described earlier) and those who place a greater emphasis on the learning of specific historical facts. The culture wars will likely continue to be an obstacle in implementing the history curriculum, but positive results on standardized tests, which tend to measure traditional content knowledge, might ease the minds of at least some critics. The next section moves beyond theoretical considerations to review research that analyzes learning outcomes associated with in-depth curriculums and authentic pedagogy.

## **Research**

### **Overview**

A wide range of studies have evaluated instructional programs and curriculums that correspond with Rossi’s definition of active, in-depth instruction. This section provides

an overview of some of the more prominent efforts to compare traditional curriculums with inquiry-based instructional approaches designed to foster critical thinking. The variables in these studies closely relate to those found in the authentic pedagogy model. This overview is followed by a more in-depth discussion of the authentic intellectual work studies. The purpose of this review is to consider the extent to which research suggests that AIW and disciplined inquiry enable students to achieve lower and higher order learning outcomes. I am also concerned with equity: does this type of instruction benefit certain students while leaving others behind? I argue that the research is inconclusive on these topics and that more authentic pedagogy studies are needed which specifically deal with social studies content.

Inquiry has been a central component of many curriculums designed to teach students to think critically in social studies. Research towards this goal has evolved over time with the major trends documented by a number of researchers (Cornbleth, 1985; Dewey, 1910; Fenton, 1967; Gross & McDonald, 1958; Hahn, 1991; Massialas & Cox, 1966; Metcalf, 1963; Newmann, 1991b; Oliver & Shaver, 1966; Parker, 1991; VanSickle & Hoge, 1991; Wallen & Travers, 1963). This section is primarily organized according to the periodization emphasized in Parker's review of literature on the promotion of critical thinking in social studies (Parker, 1991).

In the early 20<sup>th</sup> century, two main approaches were used to teach critical thinking. The first approach involved breaking down the components of critical thinking into subskills to be taught directly (Parker, 1991). Studies of this nature during the interwar period focused on propaganda resistance. During WWII, these gave way to studies designed to test the efficacy of various teaching strategies designed to help

students apply specific rules of logic (Chenoweth, 1953; Glaser, 1941; Henderson, 1958; Hiram, 1957; Rothstein, 1960).

The other dominant approach was referred to as progressive education. Progressive educators believed critical thinking could be fostered in classroom settings that permitted “a greater degree of self-determination, flexibility of curriculum, and freedom of behavior” (Wallen & Travers, 1963, p. 484). Research projects focused on the effects of specific inquiry teaching methods such as the problems-approach (Bayles, 1956; Kight & Mickelson, 1949; Quillen & Hanna, 1948) or interventions designed to evaluate classes which were more student-centered (Barratt, 1964; Elias, 1958; Rehage, 1951). Major research initiatives such as the Eight Year Study evaluated the effectiveness of a variety of progressive reforms (Aikin, 1942; Dimond, 1948; Lipka et al., 1998; Peters, 1948) .

The next major period of innovation took place during the 1960s and was influenced by the cognitive revolution in psychology. Research efforts centered on teaching students how to engage in disciplined-inquiry (Bruner, 1960; Taba, 1966) or strategies for investigating value conflicts through issues-centered curriculums (Levin, Newmann, & Oliver, 1969; Massialas, 1963; Newmann & Oliver, 1970; Oliver & Shaver, 1966). The first category included large curriculum projects whose purpose was “to shape the mindset of a generation into rational structuralist and scientific ways of seeing, and away from moral questions, social issues and social problems” (Evans, 2004, p. 129). These projects tended to focus heavily on fielding new curriculums and not as much on comparing learning outcomes with traditional instruction. One exception was the work of Hilda Taba (1964/1966) which successfully developed and tested a



curriculum designed to promote critical thinking in elementary social studies students. Taba found that a sequential curriculum which embedded instruction on critical thinking within disciplinary based inquiry lessons was able to significantly improve student thinking as indicated by their performance on the Social Science Inference Test (Taba, Levine, & Elzey, 1964). In addition, their ability to learn traditional content knowledge was not compromised (Taba, Levine, & Elzey, 1964).

Dissertations focused on this same “structure of the disciplines” inquiry model were often more evaluative (Armstrong, 1970; Dodge, 1966; Frankville, 1969; Hunkin, 1967; Madden, 1970; Rose, 1970; Williamson, 1966; Womack, 1969; Yost, 1972). In nearly every study where an inquiry model was compared with a traditional instructional approach, students in the inquiry-oriented groups did as well or better on conventional achievement tests (Armstrong, 1970; Dodge, 1966; Frankville, 1969; Hunkin, 1967; Rose, 1970; Womack, 1969; Yost, 1972). The inquiry groups also showed the greatest improvement when critical thinking or problem-solving variables were measured (Armstrong, 1970; Dodge, 1966; Yost, 1972). Two studies from this period investigated the effects of in-depth curriculums compared with coverage-based instructional programs (Johnson, 1961; Williamson, 1966) and found the in-depth curriculums were as effective in preparing students for conventional tests. Studies in later decades in geography, economics, and U.S. history (Byungro, 1991; Harmon, 2006; Mackenzie & White, 1982) also supported the use of active, inquiry-based instruction. The only exception that I encountered was a study conducted by Williams (1981) which generally found the traditional curriculum to be superior. In this study, an experimental group of 51 students who received an inquiry-based curriculum was compared with a control group of 53

students. The experimental group demonstrated significantly greater achievement on the Cooperative Topical Tests (CTTAH) for U.S. History (Williams, 1982).

Issues-centered curriculums during the 1960s built off the problem-based research of the progressive era (Hahn, 1991). The issues-centered studies that I reviewed spanned several decades and suggested that inquiry based instructional approaches do not harm the ability of students to learn factual content (Cousins, 1962; Cox, 1961; Elsmere, 1961; Gallagher & Stepien, 1996; Lambert, 1980; Lee, 1967; Massialas, 1961; Saye & Brush, 1999a). The ability of these programs to help students think critically and achieve higher order outcomes was mixed. In some cases, students in the experimental groups made significant improvements on standardized tests purported to measure critical thinking outcomes (Cousins, 1962; Lambert, 1980; Lee, 1967). Other studies failed to note noticeable differences between the control and experimental groups (Cox, 1961; Massialas, 1961). Some researchers concluded that significant advances in critical thinking, not captured on traditional tests, were still taking place based on qualitative analyses of classroom discussions (Elsmere, 1963; Massialas, 1961).

A further advance in the research on inquiry and the fostering of critical thinking in social studies took place during the 1980s as Fred Newmann worked to create a general framework for promoting higher order thinking that would be widely accepted by both researchers and teachers (Newmann, 1991a, 1991b). The design of his framework was grounded in a thorough review of research across subject areas and synthesized findings from the issues-centered and discipline-based inquiry traditions. Newmann conceived of higher order thinking as involving “a challenge that requires the person to go beyond the information given; that is to interpret, analyze, or manipulate information

because a question or a problem to be solved cannot be resolved through the routine application of previously learned knowledge” (Newmann, 1991b, p. 385). Success with these “novel” challenges involved the integration of knowledge, skills, and dispositions (Newmann, 1991b, p. 385). In order to promote these components of higher order thinking, he devised standards which eventually evolved into the authentic intellectual work model. Research associated with this model will be discussed later in this chapter.

This brief overview describes the evolution of curriculums designed to teach students to think critically in social studies courses. The degree of correspondence between these studies and the authentic intellectual work model varies. Some of the early work does not appear to have much in common with the AIW research. The narrow skills based conception of critical thinking (i.e. Henderson, 1958) isn’t very compatible with Newmann’s definition of higher order thinking or the constructivist orientation of authentic pedagogy. The research on student-centered reforms evaluated during the progressive movement is also suspect in the sense that it is difficult to evaluate the actual intellectual demands that were placed on students. On the other hand, some studies, like Chenoweth’s, had a stronger connection to Newmann’s vision of authentic intellectual work. Chenoweth’s study explicitly established a connection between a problem-based curriculum and contemporary issues while also requiring students to take action beyond the classroom (Chenoweth, 1953, p. 21). This type of curriculum would likely score high on the “connectedness to the real world” standard.

Generally speaking, the early period of experimentation (1920’s-1950s) yielded some findings to suggest that a more student-centered, problem-based approach to instruction can improve student performance. Most studies indicated that instructional

programs designed to encourage critical thinking did not harm students in their ability to achieve on conventional tests (Aikin, 1942; Barratt, 1964; Bayles, 1956; Elias, 1958; Kight & Mickelson, 1949; Peters, 1948; Rehage, 1951; Rothstein, 1960). However, the Stanford Social Education Project found that juniors in the experimental “problems” group learned less factual content about American history than those in the control who experienced a chronological curriculum (Quillen & Hanna, 1948, p. 174). Early research also indicated that it was possible to directly teach specific critical thinking skills and that students were not as likely to learn these skills through a traditional didactic curriculum (Chenoweth, 1953; Glaser, 1941; Henderson, 1958; Kight & Mickelson, 1949; Quillen & Hanna, 1948; Rothstein, 1960).

Most scholars who have reviewed the social studies literature previously described have noted that despite years of study, a coherent research base is lacking. This is due to a number of factors such as the use of diverse terminology (i.e. concept-generalization method, problems-approach, jurisprudential approach, reflective inquiry, etc.), studies grounded in differing assumptions of the nature of thinking, poor research design, and the general difficulty of implementing inquiry-based curriculums (Hahn, 1991; Metcalf, 1963; Newmann, 1991b; Onosko, 1991; Taba, 1966). Without the use of a consistent underlying theoretical framework it becomes difficult to establish significant findings among the diverse studies. Which inquiry approach is most likely to provide the type of learning outcomes (lower and higher) needed in today’s high-stakes testing environment?

The biggest factor to consider when comparing the body of research in social studies with the authentic intellectual work research is the variable of instructional

quality. Instructional methods tend to be compared without considering the intellectual challenge they represent for students (Metcalf, 1963; Newmann, Bryk, & Nagaoka, 2001; Quillen & Hanna, 1948). An analysis of classroom dialogue and the demands placed on students is important because without this analysis it is difficult to compare the level of intellectual challenge students experienced in different inquiry oriented environments. It is also possible that instruction might not have differed substantially between the experimental and control groups within some studies.

The research base also includes a disproportionate number of dissertations. These often included a number of limitations such as small sample sizes, relatively short interventions, and settings that were probably not very diverse. It was fairly common for the researcher to serve as the teacher for both the control and experimental classes (Barratt, 1964; Cousins, 1962; Massialas, 1961; Yost, 1972). While this mitigated concerns about teacher personality influencing outcomes, it presented new problems. To what extent was the teacher genuinely able to switch from one instructional approach to the other during the course of the day? Was the teacher unconsciously biased towards the experimental group?

Some dissertations also evaluated student content knowledge based on a teacher made unit test (Barratt, 1964; Elias, 1958; Lee, 1967). Today's graduation exams are more ambitious in their demands since they usually encompass an entire semester's worth of material. The consequences of sacrificing coverage for depth could be more severe for students when they are held accountable for material that was either omitted or rapidly covered between inquiry units. Schools are also held accountable for the performance of *all* of their students. While some studies evaluated the effectiveness of inquiry-based

instructional strategies based on gender, most did do not provide information regarding how other subgroups (ethnicities) performed on the outcome measures.

The strongest studies, including the Taba experiments and Indiana Experiments in Inquiry, tested a clear theoretical model, with a relatively large sample, over an extended period of time. Findings from these studies tend to favor an inquiry approach over traditional instruction in producing lower and higher order outcomes. More replication and longitudinal studies are needed to confirm their findings.

The next section provides a more in-depth look at several studies that evaluate instructional models with strong connections to authentic intellectual work. The first group of studies involves the jurisprudential model associated with the Harvard Social Studies project of the 1960s. Two additional survey-based studies provide evidence of the effects of various types of instruction on student performance. These studies are frequently cited by proponents of authentic intellectual work and are therefore most in need of examination.

### **Harvard Social Studies Project**

The Harvard Social Studies project tested a jurisprudential inquiry model which had students consider recurring public policy issues that often have no easy solution (Oliver & Shaver, 1966). The model primarily used a discussion format to get students to clarify important facts, definitional issues, and ethical considerations associated with a persistent issue (Oliver & Shaver, 1966). Phase I was a four year study conducted at the junior high (7<sup>th</sup> & 8<sup>th</sup> grade) level. The experimental curriculum was implemented by four teachers who were also researchers from Harvard.

The assessment that best evaluated the experimental curriculum was the Social Issues Analysis Test (SIAT). The SIAT included an argument analysis test, argument description and rebuttal test, oral argument analysis test, and the analytic category system (ANCAS) test which featured interview and student-led discussion components. The experimental classes outperformed the control groups on each of these tests. The results of these tests led the researchers to conclude that students could be taught to think abstractly using the jurisprudential model, especially when considering relatively simple cases (Oliver & Shaver, 1966, p. 272). Oliver and Shaver also measured student attainment of factual content knowledge and concluded that the experimental curriculum did not put students at a disadvantage when compared to students in a more conventional setting. In fact, students from the experimental group were better able to retain the factual information they learned.

The high school component of the project began in 1964. Two classes received instruction on the experimental curriculum from project staff for three years beginning in the tenth grade. The students in these classes were compared with three other groups during their senior year. The first comparison group included students at the same high school who received the experimental curriculum from regular (non-project) teachers. The other two groups, an honors group and “standard” track group, came from an affluent and academically strong school in the neighborhood (Levin, Newmann, & Oliver, 1969, p. 115).

During the final evaluation, the participating students took a variety of assessments. Among the written tests that featured lower order content were a standardized Problems of Democracy (POD) test and an open ended American History

Factual Recall Test. On the POD test, the honors control group scored the highest. The project group did as well as the other two control groups. On the open ended factual recall test, the project groups scored significantly lower than both control groups (Levin, Newmann, & Oliver, 1969). The researchers suggested that in affluent schools, where students and parents highly value education and equate success with testing, that students are more likely to be motivated and excel on conventional measures of achievement (p. 174). However, the intellectual challenge students actually experienced from the chronological curriculum at the affluent school was not investigated.

A strength associated with this research is the fact that the Harvard researchers compared the learning outcomes of the project students with students from a more academically oriented school. This lends significance to the results from the Problems of Democracy test; the closest thing to a traditional standardized assessment. However, the research still has several important weaknesses. First, the researchers did not implement pre-tests or conduct periodic assessments to measure learning outcomes until the very end of the program. There is no way to determine how much students learned as a result of the experimental curriculum. Second, a variety of curriculum innovations were implemented during the three years associated with this project. It is impossible to isolate which variables might have contributed to student success on the POD test and the other outcome measures (Levin, Newmann, & Oliver, 1969, p. 112). Finally, the general setting of the Harvard Social Studies Project was a middle class suburb. Limited data is provided regarding the ethnic diversity of the schools. More information is needed regarding whether this curriculum is effective for different groups of students.



Two later studies also evaluated curriculums based on a similar public issues discussion model as the Harvard Social Studies Project. The first one addressed criticisms levied against inquiry-based curriculums which argued that they were too advanced for disadvantaged students or slow learners. Curtis and Shaver (1980) found that slow learners can effectively engage in complex reasoning of social issues when appropriate scaffolding of materials is provided. Another study featured research on the effectiveness of a Channel 1 television segment called *You Decide* (Johnston, Anderman, Milne, Klenk, & Harris, 1994). Students in the experimental groups performed better on a test of factual knowledge related to the news events they had watched.

### **Survey-Based Research**

A fairly large study was conducted by Smith & Niemi (2001) which investigated a number of factors that might influence achievement in history. One area in particular was the impact of instructional methods on achievement. The study incorporated an ex post facto analysis of data from the 12<sup>th</sup> grade National Assessment of Educational Progress (NAEP) history exam. Data was derived from a sample of 4,465 students. A questionnaire associated with this test provided a description of the types of instruction students reported experiencing in their social studies classes. The researchers looked at the extent to which high school history courses emphasized writing complexity, reading complexity, use of alternative sources, and student discussion/debate.

The outcome measure was student performance on the NAEP, a test featuring a mixture of lower order items and items that require more extended responses and some higher-order thinking. The students who reported experiencing higher degrees of active instruction that required complex reading, writing, and discussion scored higher on this

assessment than their peers. The researchers concluded “if left with a choice of only one ‘solution’ to raise history scores, it is clear that instructional changes have the most powerful relationship to student performance” (Smith & Niemi, 2001, p. 38). This finding is promising since the variables in this study closely fit the authentic pedagogy model. However, the self reported nature of the data is a limitation. This limitation can best be seen when looking at how the researchers defined discussion. This variable mainly focused on the amount of discussion related to a specific context (i.e. whole class, small group, presentations). We have no way of determining whether the active discussion students reported experiencing was the type envisioned by Newmann; especially with indications that students and teachers seem to have very different views of quality discussions when compared with researchers (Hess, 2008a).

Another study conducted in 2001 by Lee, Smith, and Newmann analyzed how different types of instruction influence student learning outcomes on conventional tests. Their study focused on Chicago elementary schools with data from grade levels 2-8. The researchers used a 1997 survey conducted by the Consortium on Chicago Public School Research to determine the extent to which teachers used didactic or interactive instruction. They also analyzed the amount of review teachers included in their curriculum. Instructional data was paired with student performance data from the Iowa Test of Basic Skills which was a measure of reading and math proficiency.

The study included data from 384 schools, over 5,000 teachers, and over 100,000 students. Three important conclusions were made by the researchers. First, students who received more interactive instruction performed better on the ITBS. They learned 5.1% more in math and 5.2% more in reading when compared to the city average. Students

who frequently received didactic instruction tended to score below the city average. Second, interactive instruction was more frequent in the lower grades and became less frequent in the upper grades. Finally, this study noted trends to suggest that low income students and students in classes with low prior achievement levels received more didactic instruction.

The drawback of this study is essentially the same as the previous one. The use of survey instruments (as opposed to direct observation) makes it difficult to understand exactly how the interactive instruction was implemented from class to class and whether it was rigorous. This study was part of a broader Annenberg research grant in Chicago that focused primarily on Authentic Intellectual Work. The AIW studies provide greater fidelity in examining the effects of intellectual challenge on achievement. It is to these studies that I now turn.

### **Learning Outcomes: Authentic Intellectual Work (AIW)**

Reforms associated with the authentic pedagogy model have been enacted domestically (i.e. Iowa, Michigan, Washington, Minnesota, Illinois) and internationally in Australia, the Netherlands, and Singapore (Koh, Kim, & Luke, 2009; Koh et al., 2005; Roelofs & Terwel, 1999). A number of research projects have analyzed the impact of authentic intellectual work on student learning. Very few of these studies focused on the relationship between authentic pedagogy and lower order achievement outcomes. Most are concerned with the extent to which authentic tasks promote complex intellectual work by students. My review of this research begins with the studies conducted by Newmann and his associates under the auspices of the Center on Organization and Restructuring of

Schools (CORS). Newmann's work primarily took place in Chicago's public schools during a series of studies beginning in the mid 1990s. Rather than present these studies chronologically, I separate them into two main categories; AIW's impact on lower order outcomes associated with standardized tests and its impact on higher-order rubric-based measures of authenticity. In each category I distinguish between social studies research and research that focuses on other subject areas. The final section describes the progression of this line of inquiry both domestically and internationally.

### **AIW and Lower Order Outcomes**

**Subjects other than Social Studies.** The AIW studies in this section are useful even though they focus on other subject areas. In reviewing this research, the first question is whether authentic intellectual work impedes student performance on standardized tests. A study that dealt with this issue was conducted by Lee, Smith, and Croninger (1997). The goal of this research was to determine the factors that most contribute to successful school restructuring. An earlier study by the same authors (1995) indicated that smaller communal schools were more effective in promoting student learning than larger schools. The 1997 follow-up study analyzed a variety of variables, including instruction, to learn more about the reasons for this finding. It focused on 9,631 seniors in 789 high schools. Data was derived from the National Educational Longitudinal Study (NELS; 1988-1992). The NELS tracked the academic progress of participating students while associating them with their respective high schools and teachers. The researchers had achievement scores and surveys at their disposal for seniors extending back to their eighth grade year.

Lee et. al. utilized the NELS survey data from teachers and students to estimate the level and distribution of authentic instruction in schools. They then linked the results of this analysis with achievement outcomes in science and math on the NELS tests. The researchers concluded “...students attending schools that are instructionally rich and incorporate active learning and in which this type of instruction is shared widely gain more in science and mathematics achievement, both early and late in high school” (Lee, Smith, & Croninger, 1997, p. 141). They also noted that achievement gains were more equitably distributed when authentic instruction was pervasive in the school.

This study offers some interesting insights, but it also has at least one limitation; the use of survey data to estimate the authentic demands of instruction instead of actually observing instruction or collecting tasks and student work. The surveys allowed the researchers to characterize instruction in broad terms as being more or less active, but do not provide a clear indication of the intellectual challenge being offered to students. For example, the survey items asked how often students use computers, use hands-on materials or models, use books other than the math textbook, participate in student-led discussion, etc. These activities and others listed in the survey can, and often do, involve little challenge. Despite this limitation, the study still provides a rough measure of the type of instruction students’ encounter and its conclusions, when taken into account with Newmann’s other research, are an important contribution to the field. Assuming that the NELS is primarily a test of lower order knowledge, this study suggests that student performance will not be harmed by the use of authentic instruction.

A later study also addressed the issue of authentic intellectual work and student performance on standardized achievement tests (Newmann, Bryk, & Nagaoka, 2001).

This study included a mixture of grade levels to correspond with the grades in which the standardized tests were customarily administered. Researchers collected data for three years (1997-1999) from a sample that included 19 schools. The schools were a representative sample of the types of public schools found in Chicago, but were actually a little more disadvantaged than the norm (Newmann, Bryk, & Nagaoka, 2001). The most pertinent information in this study related to performance data collected on the 1400 eighth grade participants.

The research team collected two typical and two challenging math and writing assignments from each participating teacher. These tasks were evaluated by outside teachers trained in the use of the AIW scoring rubrics. Based on this analysis, the participating teachers were ranked according to the intellectual challenge represented by the tasks they submitted. The researchers compiled the scores of the students in the participating teachers' classes on the Iowa Test of Basic Skills (ITBS) and the Illinois Goal Assessment Program (IGAP).

The researchers found that students in classes that received high quality assignments scored 20% higher than the national average on the writing and math portions of the ITBS. In comparison, students who received assignments that were less demanding scored 25% less than the national average in reading and 22% less in math on average. The same basic trend was evident for IGAP scores. Students who received high quality assignments were likely to outperform their peers on the IGAP reading portion by 32 points, the math portion by 48 points, and the writing rubric by 2.3 points (Newmann, Bryk, & Nagaoka, 2001, p. 25). This study is significant because it most clearly

demonstrates the relationship between authentic instruction and achievement on lower order conventional tests.

In a third study which featured content other than social studies and focused on standardized testing outcomes, D'Agostino sought to determine the impact of instruction on Title I programs seeking to improve the reading and math achievement levels of disadvantaged students. Researchers analyzed instruction in 53 third grade Title I classrooms in 29 schools in the Chicago public school system. These schools were in high poverty areas and the majority had student populations that were 90% African American (D'Agostino, 1996). Instruction was rated based on Newmann's authentic intellectual work principles. D'Agostino found that most classrooms did not heavily emphasize AIW. In general, the math lessons tended to score higher than reading. Disadvantaged students in these schools often did not engage in lessons that featured higher order thinking related to situations they were likely to encounter in their lives outside of school (D'Agostino, 1996).

Student achievement in this study was based on specific reading and math subtests of the Iowa Test of Basic Skills that measured both higher order and lower order knowledge and skills. D'Agostino found that authentic instruction had no relation to vocabulary achievement in reading (D'Agostino, 1996). However, a moderate amount of authentic instruction was shown to improve achievement on the reading comprehension section of the test. The results for math instruction were more consistent. Students who received higher levels of authentic instruction demonstrated greater adjusted gains (pre vs. post) on both the higher order and basic skills portions of the ITBS than their peers in less authentic settings.

The studies by Lee et. al., Newmann, and D'Agostino suggest that students who experience higher levels of authentic pedagogy are not likely to perform any worse on conventional standardized tests than students in less authentic settings. The findings are strongest for elementary students in math, reading, and writing. In general, authentic pedagogy was not very prevalent among the teachers in these studies. However, when teachers did provide authentic instruction, the benefits appeared to be equitably distributed. In Newmann's study in particular, gains on the ITBS were in some instances larger for students with lower levels of prior achievement than those of their higher achieving peers (Newmann, Bryk, & Nagaoka, 2001). D'Agostino's findings were less decisive, but still suggest that lower achieving students might benefit from AIW; especially in math. The findings from these studies should be viewed tentatively, especially since two of the three focused exclusively on Chicago's public schools.

**Social Studies.** Only one AIW study measured the impact of authentic intellectual work on a lower order measure of student learning in social studies. Avery's study (1999) involved five U.S. history teachers in one urban high school in Minnesota. The teachers implemented a four-week unit on immigration where they used the same authentic essay task as the culminating activity. The students also completed a conventional 10 item multiple choice test. Avery controlled for factors that might influence student achievement to include sex, race/ethnicity, socioeconomic status, and student engagement. The main goal of this study was to determine how the level of authentic instruction would impact student performance on a common task that met Newmann's requirements for being authentic. Each teacher taught a similar lesson to prepare their students for the task (in terms of content), but the approach they used varied



significantly. Raters evaluated the classroom instruction and assigned scores based on Newmann's instruction rubric. Avery found that authenticity of instruction accounted for 40% of the differences in student performance on the task (Avery, 1999). Students who received higher quality instruction performed better on the authentic task. Avery noted a small statistical link between the level of authenticity of instruction and student scores on the multiple choice test.

### **AIW and Higher Order Outcomes**

**Subjects other than Social Studies.** Two of Newmann's studies preceded the 2001 study discussed in the previous section (Newmann, Bryk, & Nagaoka, 2001) and focused on higher order authentic outcomes. Newmann and his associates conducted a one year study on AIW in restructured schools (Newmann & Associates, 1996; Newmann, Marks, & Gamoran, 1996) and another study in 1998 designed to collect baseline data for later research efforts (Newmann, Lopez, & Bryk, 1998). I will review these studies out of chronological order since the 1998 study did not include any social studies content.

Newmann, Lopez, & Bryk collected data from math and writing teachers in grades 3, 6, and 8. Twelve schools were included in this study. They were atypical of Chicago schools in general in that the students had lower test scores and were more disadvantaged than their peers. The purpose of the study was to collect information regarding the authenticity of assignments provided by teachers in the study schools and to analyze the link between these assignments and quality student learning outcomes. The researchers gathered data from two teachers in each grade, for each subject, in each participating school. They collected four tasks (two typical, two challenging) along with

student work associated with the challenging tasks. They received tasks from 74 teachers and work from 700 students. In this study, classroom instruction was not rated.

The degree of challenge represented by the assignments was broken down into four categories: extensive, moderate, minimal, or none. The researchers noted that in all three grades, the majority of the writing and math assignments fell into the lowest two categories (Newmann, Lopez, & Bryk, 1998). The challenging assignments did tend to rate higher than the typical assignments. The writing assignments were generally more demanding than the math assignments.

Students in the classrooms that offered more authentic assignments produced work that was on average 46 percentile points higher than peers in less authentic classes (Newmann, Lopez, & Bryk, 1998). This study supports the strong relationship between assignment quality and student work while noting that quality instruction is also needed to ensure student success. This study did not attempt to control for other factors that might contribute to quality student work. It only demonstrated that a relationship exists between quality tasks and quality student work. As a baseline study, it showed that at least some students in the study schools in Chicago had the opportunity to produce authentic work and they were often able to do it (Newmann, Lopez, & Bryk, 1998).

**Social Studies.** Other AIW studies focused more attention on the link between authentic instruction and authentic achievement. A study conducted by Newmann, Marks, and Gamoran in 1996 analyzed 24 significantly restructured public schools (Newmann, Marks, & Gamoran, 1996). The social studies sample from this study (grades 9 and 10) included 23 teachers with 348 students. Newmann and his associates analyzed instruction, tasks, and student work to determine the extent to which teachers

offered authentic pedagogy, which students were most likely to experience it, and its impact on student performance. Several important outcomes were determined from this study. First, the upper levels of the AIW standards were difficult to achieve. Both the teacher authentic pedagogy scores and student work scores were, on average, below the midpoint of the range of possible scores (Newmann, Marks, & Gamoran, 1996). In testing the connection between authentic pedagogy and authentic student performance, the researchers found the level of authentic pedagogy to be the most significant predictor of quality student performance. Researchers concluded that "...an average student would increase from about the thirteenth percentile to about the sixtieth percentile as a result of experiencing high versus low authentic pedagogy" (Newmann & Associates, 1996, p. 58). A final set of findings for this study relates to equity. Newmann's analysis found that student characteristics (gender, race, ethnicity, SES) did not play a significant role in determining whether they received authentic pedagogy in the restructured schools (Newmann, Marks, & Gamoran, 1996). In addition, the effect of authentic pedagogy also seemed to be positive for all students. The only area where the effect seemed to differ was in terms of student prior achievement levels. Students with high prior achievement levels benefited more than their peers when they experienced higher levels of authentic classroom instruction (Newmann, Marks, & Gamoran, 1996). Newmann also sought to determine whether the high scoring student work indicated a bias towards any particular subgroup. According to his analysis, "Hispanics and low-SES students did not score significantly lower than whites or high-SES students, respectively..." (Newmann, Marks, & Gamoran, 1996, p. 303-304). Some achievement gaps were noted; blacks scored lower than whites and girls outperformed the boys. These achievement gaps were not

significantly greater than the gaps found on the traditional NAEP assessment (Newmann, Marks, & Gamoran, 1996).

A limitation in the study design was the fact that students who received inauthentic tasks were not really afforded the opportunity to produce quality student work (Newmann, Marks, & Gamoran, 1996). Noel criticized the study for having low inter-rater reliability among researchers evaluating student work (.54), the lack of validity data on the measurement scale, and for failing to use the same assessment tests to evaluate student learning (Noel, 1996). Other researchers have also questioned the authentic pedagogy terminology and its validity (Cizek, 1991a, 1991b; Terwilliger, 1997, 1998).

King, Schroeder, and Chawszczewski (2001) examined the impact of authentic instruction on students with disabilities in inclusion classrooms. Specifically, they looked at secondary schools with inclusionary practices and asked “to what extent are teacher-designed assessments authentic?” and “how do students with and without disabilities perform on these assessments? (p. 1).” The study included a variety of subjects (language, science, math, social studies) and grades (9/10, 11/12). The researchers collected a task and student work from teachers in two different data sets during the 1999-2000 school year. The first data set included 16 teachers from two schools. In this data set, the researchers collected and analyzed student work for the entire class pertaining to the submitted task. The second data set included 35 teachers from three schools representing the same subject areas and grades. The teachers submitted work for two students; one regular and one with a disability. This data was used for comparison purposes. The tasks in both data sets were analyzed based on the

writing and math AIW task rubrics. The student work was analyzed using subject specific AIW rubrics.

The findings for data set one indicated that the majority of the tasks failed to offer problems that showed a connection to students' lives. Nevertheless, a significant relationship was noted between the level of authenticity on the task and the quality of student work as measured by the AIW student work rubrics. The most interesting finding shows that special education students who were assigned higher quality tasks produced work of higher quality than special education students who received less authentic assignments. The difference was not statistically significant (King et. al, 2001).

In data set 2, accommodations granted to special education students were factored in to the analysis of task quality. A statistically significant amount of the tasks scored lower in authenticity when accommodations were factored in, but the researchers noted that most tasks (85.7%) scored the same. When the work produced by the matched pairs (one disabled, one regular) was analyzed, King found that 62% of the work produced by the disabled student was of equal, or higher, quality than their non-disabled peer (King et. al, 2001). Like data set 1, there was a high correlation ( $r = .68$ ) between the authenticity of a task and the authenticity of student work. King et al. concluded "Teachers who use more authentic assessments elicit more authentic work from students with and without disabilities (King et al., 2001, p. 12)."

Table 1

*Summary of Results from Authentic Intellectual Work Studies*

Study	Subject(s)	Conventional Outcomes	Authentic/Higher Order Outcomes	Equity
Lee, Smith, & Croninger (1995/1997)	Science and Math (8-12)	Students in schools with high levels and distributions of authentic instruction achieved larger gains on the NELS tests. (p. 141)	N/A	Learning is more equitably distributed when authentic instruction is “pervasive” in schools. (p. 141)
Newmann, Marks, & Gamoran (1996)	Math and Social Studies (all levels)	N/A	Authentic pedagogy in both subjects was the highest predictor of complex intellectual student work.	Authentic instruction was beneficial for all students regardless of gender, ethnicity, race, or SES.
Newmann, Lopez, & Bryk (1998)	Math and Writing (3,6,8)	N/A	“Students in the classrooms that offered more authentic assignments produced work that was on average 46 percentile points higher than peers in less authentic classes” (p. 39)	Students in this sample were more disadvantaged than others in Chicago.
Avery (1999)	U.S. History	Students not harmed by authentic instruction on conventional 10-item test.	Students who experienced authentic instruction performed better than peers on authentic, higher-order essay measure	
Newmann, Bryk, & Nagaoka (2001)	Math and Writing (3,6,8)	Authentic instruction enabled students to perform at a higher level on the ITBS and IGAP.	N/A	Students in this sample were generally more disadvantaged than their peers in other Chicago schools.

Table 2

*Summary of AIW Studies with an Explicit Focus on Disadvantaged Students*

Study	Focus	Subject(s)	Conventional Outcomes	Authentic/Higher Order Outcomes
D'Agostino, 1996	Title I and Low SES	Math and Reading (3)	Higher levels of authentic instruction positively correlated with improved math scores <sup>1</sup> . Moderate use of authentic instruction appears to best promote reading comprehension.	
King, Schroeder, & Chawszczewski, 2001	Special Education Students	Language, Math, Science, Social Studies (9-12)		Special Education students who received high levels of authentic pedagogy achieved at higher levels than regular students who received low levels of authentic pedagogy.
Amosa et al., 2007	Aboriginal and Torres Islander; Low SES (Australia)	Not Indicated	No conventional measure	Indigenous students who received high quality tasks produced work that on average exceeded work produced by non-indigenous students who received low quality tasks. Low SES performed better than high SES when both given high scoring tasks.

*Note:* Outcome measure is the ITBS. Amosa's study used a construct unique to Australia which incorporated AIW

## **Gates Foundation Research**

The High School Grants Initiative supported by the Gates Foundation adopted the authentic intellectual work model as a way to evaluate the effectiveness of its initiative to redesign or build new high schools based on a small learning community model. In 2002/03, researchers collected and evaluated tasks and student work as a pilot study with teachers in Washington State (AIR/SRI, 2004). During the next year, the program was implemented nationwide and the AIW rubrics were used as one measure to compare redesigned schools with traditional ones (AIR/SRI, 2006). A final study completed in 2007 evaluated the performance of Foundation schools with the baseline data collected from the pilot study (AIR/SRI, 2007). In each of these studies, quality student work was defined by the criteria specified in Newmann's AIW framework and therefore encompassed higher order outcomes such as construction of new knowledge in English/Language Arts and reasoning and problem-solving in mathematics (AIR/SRI, 2006).

The 2006 study included limited analyses of the relationship between authentic pedagogy and learning outcomes; both standardized and authentic. After controlling for a number of factors, researchers in this study found a significant positive relationship between quality student work in English/Language Arts (ELA) and improved standardized test scores in reading. In mathematics, the relationship was positive, but not statistically significant (AIR/SRI, 2006, 2007). In analyzing the relationship between authentic assignments and higher order authentic outcomes, the researchers focused on



elements of the AIW framework individually (assignment rigor & relevance) to more accurately pinpoint the elements of an authentic task that most influence student learning. This makes it difficult to make direct comparisons between the Gates' studies and other studies conducted by Newmann and his associates. However, some general conclusions still apply. In both subjects, authentic assignments were positively associated with quality student work (AIR/SRI, 2006, 2007). ELA students responded to challenging tasks by producing work of better quality than students in the comparison schools (AIR/SRI, 2006). However, in math, where the assignments from the Foundation schools were slightly better than comparison schools, the student work did not exceed that of the traditional schools. The researchers noted that most of the math assignments did not score very well from either type of school and that teachers may be encountering difficulties in implementing constructivist assignments.

The other Gates Foundation study with implications for my research focused on the redesigned schools in Washington State and evaluated instructional changes and learning outcomes compared to baseline data collected in the pilot study (AIR/SRI, 2007). In conducting their analysis, the researchers controlled for student demographics, prior achievement, teacher characteristics, and several other variables that potentially could influence achievement. Student performance was based on the Washington State 10<sup>th</sup> grade achievement tests (WASL) and in this instance researchers noted a positive relationship between quality student work and math scores (AIR/SRI, 2007). There was not a positive relationship for language arts. The researchers, noting the discrepancy of outcomes between this study and the previous one, hypothesized that student work

quality might have “higher correlations with tests other than the WASL” (AIR/SRI, 2007, p. 21). This has implications for my study in that authentic pedagogy may succeed in enabling students to perform well on certain types of standardized tests, but not others.

This study also analyzed the extent to which authentic tasks result in complex intellectual work in student products. As in the previous study, a high correlation was noted between authentic tasks and quality student work for ELA and math (AIR/SRI, 2007). Tables 3 and 4 summarize the learning outcomes associated with authentic pedagogy as established by the Gates Foundation research.

Several important points should be made about these studies. First, an explicit goal of the Gates Foundation was to target disadvantaged student populations. In an analysis of whether the Foundation was meeting this objective, the ARI/SRI 2006 report indicated that “Two thirds of new schools and almost 80% of redesigned schools exceeded their district averages for enrollment of students eligible for free or reduced-price lunch and or enrollment of students from minority backgrounds” (p.16). The outcomes described in these reports show that authentic pedagogy can achieve at least some success with students from disadvantaged backgrounds.

Another important aspect of these studies is their unique methodology. As previously noted, assignment rigor and relevance were analyzed independently as variables instead of together as part of a composite authentic task variable. The relevance variable included elements of Newmann’s connection to students’ lives standard while the rigor variable examined the extent to which the task required construction of knowledge. In analyzing the tasks provided by teachers, the researchers noted low levels

of relevance in general even though the restructured schools achieved better scores than the more traditional schools. Assignment relevance and rigor were strongly correlated. When examining the impact of the two variables on student work quality, the rigor variable appeared to have the more direct, positive impact on student performance. The researchers argued that future research should stick with their methodology to more precisely understand how assignment rigor and relevance influence student performance (AIR/SRI, 2007).

These studies also demonstrate the difficulty of the AIW scales. Even after the Foundation schools were redesigned, most math assignments were rated as showing little to no rigor (AIR/SRI, 2007). This was also the case when math tasks from traditional schools were evaluated (AIR/SRI, 2006). ELA assignments rated a little better, but 71% still fell below the substantial rigor level (AIR/SRI, 2007). This reinforces the need for standards of intellectual quality to improve the way constructivist practices are implemented in the classroom.

Finally, the importance of observational data is underscored by these studies. While observations were conducted to note qualitative trends, the AIW instruction rubrics were not used in any of the analyses. The researchers attributed lower than expected performance gains in math to teachers adopting more rigorous assignments without corresponding improvements in their instruction (AIR/SRI, 2007). A recommendation from this study is to link the analysis of classroom instruction with assignments in order to determine how they work together to influence student work quality (AIR/SRI, 2007). This was also a recommendation provided by Bruce King, a researcher affiliated with

Newmann's studies (B. King, personal communication, Nov. 21, 2007). I incorporated this recommendation into the design of this dissertation study.

Table 3

*Gates Foundation Studies: Authentic Student Work and Performance on Standardized Tests*

Study	Purpose of Study	Subject Focus	Sample	Data Collected	Outcome Measures	Results
AIR/SRI 2006	Compare 12 new Foundation high schools in 8 regions across the U.S. with 8 traditional schools	English/ Language Arts (ELA) and Math	ELA: 113 students, 16 teachers, 8 schools  Math: 92 students, 20 teachers, 8 schools	Tasks, Student Work	Scores from multiple state achievement tests converted to common metric based on norms from the CAT-6 and SAT-9  ELA work relates to 10 <sup>th</sup> grade reading test scores	ELA (+)* Math (+)
AIR/SRI 2007	Gauge improvement for 12 Foundation Schools from pilot study	Same as above	ELA: 71 teachers, Math: 68 teachers	Tasks, Student Work	10 <sup>th</sup> grade Washington State achievement tests (WASL)	ELA (-) Math (+)*

*Note.* (+)\* = Significant Positive Relationship; (+) = Positive Relationship, but not significant at .05 level; (-) = No relationship.

Table 4

*Gates Foundation Studies: Relation between Authentic Tasks and Student Work*

Study	Purpose of Study	Subject Focus	Sample	Data Collected	Outcome Measure	Results
AIR/SRI 2006	Compare New Foundation Schools with Traditional Schools across country	English/ Language Arts (ELA) and Math	ELA: 89 teachers Math: 81 teachers	ELA: 717 tasks & associated student work  Math: 606 tasks & associated student work	AIW rubrics	Authentic tasks were closely associated with quality work in ELA (p. 38). Assignment rigor is more closely associated with quality work than relevance in math (p. 48).
AIR/SRI 2007	Gauge improvement for 12 Foundation Schools before and after redesign	Same as Above	ELA: 71 teachers Math: 68 teachers	ELA: 509 tasks, 966 student responses  Math: 523 tasks, 1078 student responses	AIW rubrics	Strong positive correlation between high scoring authentic tasks and student work quality in both subjects (p. 18).

## **International Research**

In Australia, two states have implemented an elaborated version of the AIW framework to improve instruction in their schools. James Ladwig, who worked with Newmann in the 1990s to help develop and test the construct, is the common link between American and Australian research related to authentic pedagogy. The Australian studies are comparable with studies in the United States since the Australian models include all of the original components of Newmann's work. In both settings (Queensland & New South Wales) researchers found authentic pedagogy to be rare in their schools (Gore, Ladwig, Lingard, & Luke, 2001; Ladwig, Smith, Gore, Amosa, & Griffiths, 2007). Studies in Singapore report a similar finding (Koh, Kim, & Luke, 2009; Koh et al., 2005).

Queensland launched a three-year school reform program called the New Basics in 2000 in specific trial schools. Thirty-eight schools were selected from the 1,296 state schools in Queensland (Education Queensland, 2004, p. 23). Trial schools included students from less advantaged backgrounds than their peers at non-trial schools (Education Queensland, 2004, p. 11). The reform project focused primarily on comparing learning outcomes of trial students with those of non-trial students.

The trial schools associated with the New Basics demonstrated a departure from regular schools in their documented use of rich authentic tasks and authentic instruction. Since the instructional program was focused on higher order outcomes, reformers did not anticipate it having much effect on basic skills development. The teachers in the trial schools administered their usual conventional tests to students. In schools that adopted the New Basics, external standardized tests (years 3, 5, 7) indicated no general decline in literacy and numeracy as compared to the rest of the state schools (Education

Queensland, 2004). The focus on authentic outcomes did not result in diminished returns on these traditional tests. A later study analyzed numeric and literacy scores for New Basics schools in 2004 and 2005. Once again, these schools showed no evidence of a decline in scores. In fact, “there is evidence that the scores of lower achieving students in New Basics schools are rising” (Lake Corporate Consulting, 2006, p. 1).

Students in both types of schools also completed two different types of higher order assessments. The first higher order assessment was the International Schools’ Assessment which focused on reading and math. The second assessment was the World Class Test; an interdisciplinary problem-solving assessment. On both of these measures, no real difference was identified between trial students and non-trial students in higher order ability (Education Queensland, 2004, pp. 38,41). The researchers viewed this as significant since trial students generally were more disadvantaged than their peers at non-trial schools.

A second Australian state enacted similar reforms. In New South Wales, authentic pedagogy was incorporated into a model known as Quality Teaching. The Systemic Implications of Pedagogy and Achievement in New South Wales Public Schools (SIPA) longitudinal study (2004-2007) provided data to document the effects of this program. The schools included in SIPA offer a representative sample of students from varying grade levels, school settings, and socioeconomic backgrounds (Ladwig, Smith, Gore, Amosa, & Griffiths, 2007). The first study based on this data was explicitly designed to replicate Newmann’s authentic pedagogy research in the Australian context (Ladwig, Smith, Gore, Amosa, & Griffiths, 2007).



In this study, Ladwig et al. evaluated instruction in grades 4 and 8. Tasks were collected primarily for Math and English. Other subjects such as Science, PDHPE (Health/PE), and an area similar to social studies called Human Society and Its Environment (HSIE) were also included in the analysis. However, only one task was from HSIE at the secondary level. These tasks (78 total) were analyzed from 26 SIPA schools. Student work came from 1,374 students. This study found a significant positive ( $p < .001$ ) relationship between high scoring tasks and quality student work even when controlling for other factors that might influence achievement (prior achievement, gender, SES, etc.).

A second study that utilized SIPA data took a closer look at learning outcomes for disadvantaged students (Amosa, Ladwig, Griffiths, & Gore, 2007). Amosa et al. studied the effects of authentic instruction on indigenous students and students from low socioeconomic backgrounds in New South Wales. As part of their study they collected 95 tasks from 121 teachers in 19 primary schools and 11 secondary schools (Amosa, Ladwig, Griffiths, & Gore, 2007). Out of the 1,912 students in the sample, 180 were Aboriginal or Torres Strait Islander. The sample was also divided according to the SES background of the students.

Amosa found that as tasks became more authentic, achievement also became more authentic for indigenous students and non-indigenous students alike. The work of indigenous students remained below non-indigenous students when aggregate comparisons were made for students who received low quality tasks. The same was true when comparisons were made for students who received high quality tasks. However, the indigenous students who received high quality tasks produced work that on average

exceeded work produced by non-indigenous students who received low quality tasks (Amosa, Ladwig, Griffiths, & Gore, 2007).

The most important finding in this research was the fact that when tasks of high intellectual quality were given to students from both low and high SES backgrounds, the students from a low socioeconomic background actually *performed better* than students from a high SES background (Amosa, Ladwig, Griffiths, & Gore, 2007, p. 6). This is the only AIW related study that has produced this finding. It is possible that the students from lower SES backgrounds were better prepared for the tasks, but this is unknown since classroom observations were not a part of the study (Amosa, Ladwig, Griffiths, & Gore, 2007).

The authentic intellectual work studies, taken in sum to include those in the United States and elsewhere, include a number of promising findings. However, these findings should be viewed tentatively. The researchers are making judgments about the intellectual demands of teachers based on limited sets of data. A variety of circumstances could cause these judgments to be flawed. Much rests on the ability to collect quality teacher data. The cooperativeness of teachers was an issue in some of these studies (Wenzel, Nagaoka, Morris, Billings, & Fendt, 2002). It is possible that some teachers merely gave researchers tasks to get them to go away or otherwise changed their routine because they were being studied. This is not a unique problem in educational research, but it should still be considered when weighing the significance of these findings, especially when they hinge on categorizing teachers based on their AIW scores.

Table 5

*Summary of International Studies*

Study	Focus	Method	Data Collected	Results
New Basics Research Report, 2004	Compare achievement of students in 18 trial schools with students in 21 non-trial schools.	Comparison of student work samples (traditional folios vs. rich tasks)	26 traditional folios and 26 rich tasks. 256 observations over three years	Student work on rich tasks perceived as rigorous by experts and community members
Based on data from Queensland School Reform Longitudinal Study (QSRLS)	Trial students received authentic instruction and completed authentic “rich” tasks  Grades 3,6, & 9	Instruction analyzed based on scored classroom observations  Student learning measured through external standardized tests & rich tasks	Conventional test results from the International Schools’ Assessment (Reading & Math) & World Class Test (inter-disciplinary problem-solving)	No decline in general literacy and numeracy as compared to state schools  No significant difference between trial students and regular students on two higher order assessments.
Ladwig, Smith, Gore, Amosa, & Griffiths, 2007	Replicate Newmann’s authentic pedagogy research in the Australian context  Grades 4 and 8	Hierarchical Linear Modeling  Analysis of challenging tasks and student work	78 tasks from 26 SIPA schools (primarily English and Math)  Student work from 1,374 students collected in 2005	Significant relationship ( $p < .001$ ) between high scoring tasks and quality student work

Table 5

*Summary of International Studies (Cont.)*

Study	Focus	Method	Data Collected	Results
Amosa, Ladwig, Griffiths, & Gore, 2007 New South Wales	Focused on disadvantaged students (indigenous and low SES students)	Hierarchical Linear Modeling (HLM) Rated intellectual quality of tasks provided by teachers  Student work analyzed using Newmann's task rubric  *No classroom observations	95 tasks from 121 teachers in 19 primary schools and 11 secondary schools  1,912 students (2,913 pieces of student work)	Indigenous students who received high quality tasks produced work that on average exceeded work produced by non-indigenous students who received low quality tasks  When tasks of high intellectual quality were given to students from both low and high SES backgrounds, the students from a low socioeconomic background actually <i>performed better</i> than students from a high SES background
Koh et. al., 2005 Singapore Pre-Intervention Study	36 Singapore Schools (18 primary, 18 secondary)  Subjects: English, Math, Science, and Social Studies	Tasks collected and scored by experienced master teachers in respective subject areas using standards consistent with AIW	Four high, medium, and low quality tasks from each teacher	Tasks were generally of low intellectual quality with the exception of primary social studies. Students were generally not afforded the opportunity to produce work that would score high on Newmann's AIW scale.

The limitations of using statistical analysis to represent an inherently complex event such as classroom instruction also should be considered. The researchers were diligent in their efforts to control for a variety of variables, but it is always possible that other factors contributed to the positive outcomes described in these studies. These could include teacher variables (personality, management style, etc.) or some of the variables included as part of the “productive pedagogies” model Ladwig worked with in Queensland.

### **Adding to the Research base**

In reviewing the major AIW studies, the need for replication in secondary social studies classrooms is evident. Few studies focused on the impact of authentic pedagogy on lower order learning outcomes. The studies that did address this relationship were not interested in social studies content. Newmann’s 2001 study provided the best evidence that authentic instruction helps students on tests of basic knowledge, but it was focused on math and writing. Several other studies provided similar results, but included little data on the nature of the conventional assessments (Lee, Smith, & Croninger, 1997; AIR/SRI, 2006/2007). It is difficult to determine whether they were as heavily weighted towards lower order knowledge as some history graduation exams. The present study is therefore important because it seeks to determine the impact of authentic pedagogy on learning using an assessment that is almost entirely fact based and has high-stakes attached to it. The addition of a higher order assessment to the study should enable me to tap a broader range of learning outcomes that may be enhanced through authentic pedagogy.

The AIW studies to date have also included a relatively limited sample of high school social studies teachers (i.e. 6 for Newmann's 1996 study). Five studies included social studies content, and most of these involved multiple subjects at different grade levels. More studies are needed to gain a better appreciation of the intellectual demands placed specifically on students in secondary history classrooms.

My study has the potential to add to the literature on authentic intellectual work because it incorporates some of the most important design modifications recommended from earlier research. I am able to more rigorously measure authentic pedagogy because the tasks supplied by participating teachers are linked to instruction. Less guess work is involved in determining the teacher's intent. In addition, I've built off of Newmann and Avery's work by providing students with a common higher order essay. This allows all students in the study to demonstrate their ability on authentic tasks which should provide a better indication of the role authentic instruction plays in promoting higher order outcomes. Since Avery's study involved a small sample, it should be replicated in different contexts with more social studies teachers. I also believe it is important to look at learning outcomes that span a semester. Avery's findings focused on one unit and suggested that authentic instruction positively impacts student performance on a higher order essay and a 10 item test. However, authentic instruction usually takes more time thus limiting the ability of teachers to cover all the necessary content on state developed standardized tests. A useful study for practitioners would seek to determine if students are able to excel on higher order tasks while also gaining the necessary content to pass a graduation exam of basic skills.

Finally, it is important to determine what works for different groups of students on the exam that arguably most determines their future academic success. Graduation exams in Alabama and elsewhere are the key accountability mechanism used by education stakeholders to gauge achievement. Students who pass the graduation exam early often have opportunities for more advanced study. Conversely, those who fail may be placed in remedial courses where they are more likely to receive drill and practice oriented instruction (Kornhaber, 2004; McNeil & Valenzuela, 2001; Oakes, 2005).

Two important factors related to equity need further study. First, who has access to authentic instruction and to what extent does high stakes testing influence its distribution? Newmann's study (1996) indicated that authentic instruction is equitably distributed, but this was in a best case scenario of restructured schools. Secondly, most prior studies indicate that all students benefit from authentic pedagogy. When dealing with high stakes social studies exams of lower order content, does this finding continue to be true? Do certain students need direct instruction to make up their deficit in content knowledge (Delpit, 1995)? These are important questions which this study attempts to investigate.

The authentic pedagogy model challenges education stakeholders to create better accountability mechanisms that have meaning and relevance in the real world. As argued by Wiggins, authentic assessments should help teachers to "improve performance, not just monitor it" and prepare students for the types of intellectual challenges they are likely to face as adults (Wiggins, 1993b, p. 5). Adoption of this model requires a commitment to establishing conditions in schools that enable educators to work

collaboratively to translate the AIW standards into effective classroom practice (Avery & Palmer, 1999; Stewart & Brendefur, 2005).

Research by Stiggins (1992) and others suggests that teachers need a good deal of support to develop assessment literacy. Teachers would need significant training in order to create and evaluate authentic tasks. The demands and challenges associated with implementing constructivist teaching are also well documented (Onosko, 1991; Rossi, 1995; Saye & Brush, 2002). The teaching force would likely face a large learning curve in adopting reforms based on the authentic pedagogy model. A change in education policy in America would require a major investment in professional development resources. This study should provide policy makers with a better basis for deciding whether to support this investment.

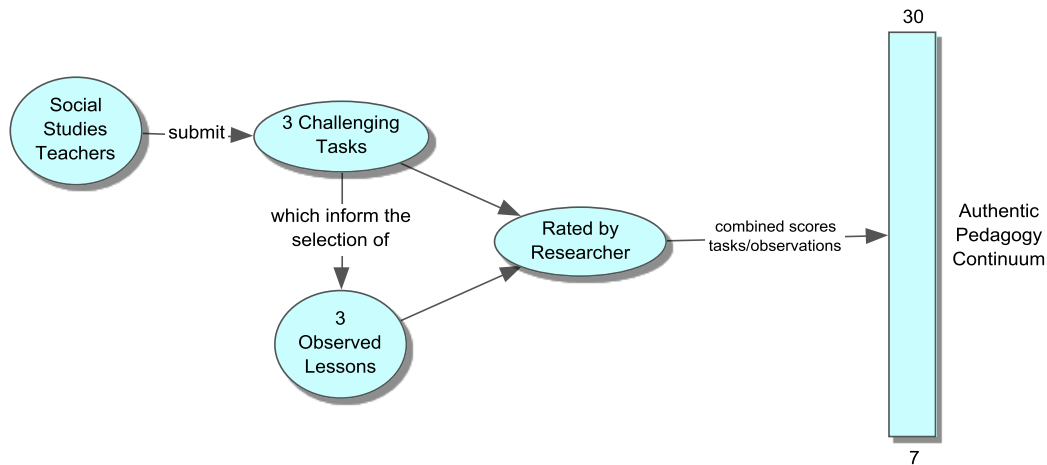


### CHAPTER THREE: METHODOLOGY

This study investigated the manner in which teaching influenced student performance, particularly on the standardized high-stakes social studies tests deemed most important by many policy makers. Instruction was analyzed in terms of its authenticity. Newmann's authentic pedagogy model provided a way to measure the extent to which instruction engaged students in activities that required construction of knowledge to solve meaningful problems that have value beyond school. Prior studies in a variety of subject areas indicate that students learn more when their teachers routinely provide authentic intellectual challenges. This research was an effort to determine the effect of authentic pedagogy in history classes.

My study used rubrics developed by Newmann and his associates that measure both instruction and assigned tasks on a numerical scale that runs from 7 to 30. Scores on the low end generally represented more teacher-centered, didactic classrooms while higher scores were associated with inquiry oriented classrooms that emphasized higher-order thinking, deep knowledge, substantive communication, and connectedness to the real world (Newmann, King, & Carmichael, 2007; Newmann, Secada, & Wehlage, 1995). I viewed this scale as a continuum reflecting the extent to which teachers provided intellectually challenging instruction. The teacher data, when paired with student scores on a lower and higher order assessment, provided the basis for determining

the impact authentic pedagogy on student performance. Figure 1 depicts this relationship.



*Figure 1. Process for determining Authentic Pedagogy Scores.*

In order to test this association, I initiated a study with a school system in Alabama beginning in January, 2008. I collected data from the 9<sup>th</sup> grade history teachers at a junior high and the 10<sup>th</sup> grade history teachers at a high school. I was able to recruit all of the social studies teachers initially assigned to these grade levels (N=8). In addition to teacher data, a variety of student data was collected by the school system. The student data specifically target tenth graders who took social studies during the 2007/08 and 2008/09 school years. The study had four main purposes: 1) to determine the extent to which social studies teachers at the study schools utilize authentic pedagogy 2) to develop a clearer understanding of how exposure to authentic pedagogy in coursework influences the ability of all students to perform at high levels on assessments that require lower and higher order knowledge 3) to determine if experiences in authentic pedagogy classrooms for multiple courses result in improved performance when compared to

students who have such experiences in only one course and 4) to determine the impact of authentic pedagogy on students from different socio-economic and ethnic backgrounds.

These purposes were translated into five research questions which are stated on page 119.

This chapter describes how the study addressed each of these purposes. The first section is a description and justification of the study design. This is followed by an overview of the study setting and participants. Once the context of the study is established, I transition into an analysis of the various research instruments and a phase by phase description of the data collection process. This sets up a concluding discussion of the data analysis procedures and study limitations.

## **Study Design**

This study was a mixed methods investigation of existing instruction at the study schools (Feilzer, 2010; Greene, 2008; Johnson, Onwuegbuzie, & Turner, 2007; Patton, 1987). Quantitative data alone (i.e. standardized test scores) does not tell us much about the types of instructional experiences that are helpful in producing desired learning outcomes. Descriptive information must also be included in the study design to determine what works for different types of students on tests like the Alabama High School Graduation Exam (AHSGE). A mixed-methods study provided greater fidelity in capturing the broad range of learning that takes place in a social studies classroom and was therefore a better approach for measuring the overall quality of instruction.

The quantitative dimension of the study required the use of statistical measures to correlate student performance data with the authentic pedagogy score of students' 9<sup>th</sup> and 10<sup>th</sup> grade social studies teachers. Instead of using an experimental "treatment", teachers (and, by extension, their classrooms) were differentiated based on pedagogy through their

placement on the authentic pedagogy scale. In analyzing the data, I sought to determine whether higher teacher scores on the authentic pedagogy continuum translated into statistically significant student achievement gains when other factors that might influence achievement were controlled.

The qualitative data, derived primarily from interviews (see Appendix A) and field notes, were converted to a numerical scale using the AIW rubric for instruction. I also conducted a content analysis of the tasks provided by teachers and converted this data in a similar fashion. This data was then subjected to inferential statistical analysis to determine achievement outcomes. The field notes provided a record of the features that distinguished higher scoring classes on the authentic pedagogy scale from lower scoring classrooms. The analysis of data from the field notes is described in chapter four.

### **Project setting and description of participants**

Central High School and Central Junior High School (pseudonyms) were selected from a school system in Alabama to participate in this study. The selection of these schools represented purposeful sampling. I chose these schools primarily because previous work in these schools gave me confidence that there would be some teachers who might be expected to score relatively high on the AIW rubrics. These schools were also easier to visit on a routine basis than alternative schools given a limited budget. Finally, the school system was selected because it was willing to provide important student achievement data that probably would not have been as easily available in other areas of the state.

The study schools are situated in a city of approximately 43,000 people. The area constitutes the most rapidly growing area of the state. Even though many people are

attracted to the area, poverty is still a problem. Fourteen percent of the families with children under 18 in the community live under the poverty line (U.S. Census Bureau, 2000a). The community is 78% white, 17% African American, and 3% Asian (U.S. Census Bureau, 2000b). These statistics do not include international students that live in the city while attending a university that is within commuting distance. They also do not accurately reflect the number of Korean families moving to the area as part of the growing automobile manufacturing industry. A recent accreditation report, conducted by the school system, noted that 42 different languages are spoken in the homes of students from the district. Overall, the area offers a relatively small town atmosphere with the economic and social benefits you might associate with a bigger city.

The high school that took part in this study has been recognized as among the best in the country by Newsweek magazine (Kantrowitz & Wingert, 2006). This designation was based on a ratio: the number of students who take Advanced Placement or International Baccalaureate exams divided by the number of seniors who graduate. The high school's enrollment in 2007 was 1,156 students while the junior high had 908 students (Central District Accreditation Guided Self Study, 2007). Despite being relatively large schools, the students at both schools enjoy some advantages that might be associated with smaller school settings. The system employs enough teachers to maintain small class sizes (average - 17 CHS, 19 CJHS). Teacher salaries rank among the top in the state enabling the schools to attract top applicants each year. The high school offers a number of challenging programs such as the International Baccalaureate Program, Advanced Placement courses, and dual enrollment. The system drop-out rate is substantially less than the state average. The vast majority of the students graduate and

pursue some form of higher education. The graduating class in recent years included numerous AP scholars, National Merit awardees, and students with GPAs over 4.0. All schools in the system have met adequate yearly progress standards for the past three years.

The schools, while atypical in some regards, were still a good choice for this study. It is true that students at Central High School typically perform above the state average on the social studies graduation exam (see Table 6). However, the scores at Central High follow the same basic trend found in other schools across the state. In 2008 and 2009, Central High students in all eligible grades (10-12) scored lower on the social studies graduation exam than any other subject (Alabama Department of Education, 2009a). Since Central’s students are struggling with the same exam as students in other schools, perhaps some useful generalizations can be made from the results of this study.

Table 6

*Comparison of Tenth Grade Graduation Exam Passage Rates*

	2008 Passage Rate	2009 Passage Rate
Central High	67%	79%
State Average	52%	62%

*Note.* Data derived from Alabama State Department of Education Accountability Reports for the 2008 and 2009 school years.

**Teachers.** As mentioned previously, social studies teachers were recruited at the ninth and tenth grade levels for this study. Tenth grade teachers were selected to participate because this is the first year when students are eligible to take the Alabama High School Graduation Exam. The eleventh grade was not a viable option because many students in this system pass the exam on their first attempt. More student achievement data was available using this sample as opposed to other alternatives. I

included the ninth grade teachers to determine the nature of the social studies instruction students received while in junior high. Research suggests that multiple years of inquiry-based instruction may have a larger impact on student performance than more limited exposure (Klentschy, Garrison, & Amaral, 2001). The 9<sup>th</sup>/10<sup>th</sup> grade design allowed me to test this finding in history classes. I was able to examine the impact of instruction over the course of multiple years on the performance of students on the Alabama High School Graduation Exam.

Once I narrowed the teacher sample to the 9<sup>th</sup> and 10<sup>th</sup> grade, I recruited all of the social studies teachers in order to maximize the potential for capturing the range of authentic pedagogy at the study schools (see Appendix B). As the study progressed, some teachers retired or had their teaching responsibilities adjusted by the administration. I did not add any new teachers to the study after the initial recruitment period. A descriptive summary of the teachers involved in the study is provided in Table 7.

The teachers proved to be an interesting sample due to the diverse range of courses they taught. The graduation exam focuses exclusively on U.S. history content. However, some teachers in this study taught World History (primarily 9<sup>th</sup> grade teachers) and two of the high school teachers taught Advanced Placement (AP) European History. In addition, the AP class sections and some U.S. history sections ran for an entire school year, while most other classes were a semester in duration. Complicating things further, some students took 9<sup>th</sup> grade World History again if they didn't pass it the previous year or perhaps if they transferred from another area and needed the credit. Data from this study enabled me to investigate whether access to authentic pedagogy was influenced by

the courses students' took and whether certain courses/course designs were more effective in promoting student achievement.

Table 7

*Descriptive Statistics for Teacher Sample*

	Junior High (4)	High School (4)	Total (8)
Percent Male	50%	100%	75%
Percent Caucasian	100%	100%	100%
Percent with advanced degree <sup>1</sup>	75%	50%	63%
<b>Age</b>			
26 to 35:	25%	50%	37.5%
36 to 45:	25%	50%	37.5%
46 to 55:	50%		25%
<b>Total Years Teaching</b>			
3 to 5:	25%		12.5%
6 to 10:		25%	12.5%
11 to 15:	25%	75%	50%
More than 16:	50%		25%

*Note:* Advanced degrees include master's degree and Ph.D.

**Students.** Students' data was gathered to determine individual learning outcomes associated with authentic pedagogy. It provided a window into whether authentic pedagogy benefited certain students more than others. Student data was also used as the basis for aggregating class level effects for statistical analysis. Data for all tenth graders at Central High School who took social studies classes during the Fall 07, Spring 08, Fall 08, and Spring 09 semesters were included in this study. This included both regular classes and AP courses. The AP European history course was open to any student willing to take the challenge although students who took this course were generally not given the option to drop if it proved to be too difficult.



An initial concern in trying to obtain a student sample for this study was the potential that the students most needed (disadvantaged, poor academic achievers) might opt out of the study. Since student performance data and demographic information was already routinely collected for analysis by the school system, the school system organized the data and coded it in order to maintain student anonymity when the dataset was sent to me. This strategy allowed me to include the data of all grade level students as part of the study (assuming the pertinent data was provided for each student). Organizing the study in this fashion maximized the potential relevance of results by including the widest possible range of students. This ensured a comparison could be made between the less advantaged students at the study schools and similar groups of students across the state. Table 8 provides a breakdown of the number of students for whom I collected data by semester and course. Students who had multiple social studies teachers during the 10<sup>th</sup> grade were excluded from the sample. This included nineteen students in 2008 and eleven students in 2009.

Table 8

*Student Participation by Course*

<b>2008</b>		<b>2009</b>	
Advanced Placement European History	99 (28.2%)	Advanced Placement European History	104 (22.9%)
U.S. History/Geography 10 (Sem.)	220 (62.7%)	U.S. History/Geography 10 (Sem.)	179 (39.4%)
U.S. History/Geography 10 Alt. (Year)	21 (6%)	U.S. History/Geography 10 Alt. (Year)	155 (34.1%)
U.S. History/Geography 10 Co-Teach (Inclusion)	11 (3.1%)	U.S. History/Geography 10 Co-Teach (Inclusion)	16 (3.5%)
Total	351	Total	454

A sample must include a certain number of students to have enough statistical power for the regression analysis. Statistical power analysis was completed as part of the planning process to determine whether a proper relationship existed between the sample size, significance criterion, population effect size, and power to prevent type I and II errors (Cohen, 1992). In my study, the desired power was .80 with a significance criterion of .05. The hypothesized effect size was medium according to the ES index for regression analysis. Determining the sample size thus required taking the number of predictor variables and multiplying by 10 (signifying ten students needed per independent variable) (Stevens, 2002). My sample of 805 students easily supports the number of independent variables needed for the study.

### **Instrumentation**

The instruments in this study served three purposes: to classify the level of authentic pedagogy used by the teachers, to determine the prior academic ability of the students, and to measure academic achievement on lower and higher order tasks. The only instruments specifically developed for this study were the higher order essay assessments. The other instruments were either state assessments or rubrics created by Fred Newmann and his associates.

**Assessing Authentic Pedagogy.** This study replicated Newmann's previous research on authentic pedagogy. As a result, I used essentially the same AIW rubrics (see Appendices C-F) to allow comparisons to be made across studies. The AIW rubrics incorporated a complex set of research based criteria into a series of different instruments for instruction, tasks, and student work making them a) tightly focused on the AIW construct and b) more efficient than alternative instruments that measure single

dimensions (i.e. higher order thinking). The fact that they have been field tested in social studies classes with students at the ninth and tenth grade levels made them ideal for use in this study. The AIW rubrics are valid instruments based on their significant construct, face, content, and predictive validity.

Construct validity is concerned with how well a researcher operationalizes theoretical ideas. The AIW framework, its associated rubrics, and other theories that form the basis of this construct are explained in detail in a number of articles and studies (Berlak et al., 1992; Newmann & Archbald, 1988; Newmann & Associates, 1996; Newmann, Secada, & Wehlage, 1995; Resnick, 1987; Wiggins, 1993a). The rubrics have been field tested extensively for over 12 years. As the rubrics have been applied in studies associated with a diverse range of academic subjects, they have been steadily revised and sharpened through dialogue with disciplinary subject matter experts and education professionals. In the process, certain stand alone dimensions of authentic intellectual work have been combined or removed altogether to help researchers make clearer distinctions between the standards being measured as part of the AIW framework. For example, the original task rubric included standards for organization of information and consideration of alternatives. Later versions of the rubric incorporated the language of these standards into a single standard called construction of knowledge. This streamlining and clarification of language over time enhanced construct validity by enabling researchers to more precisely describe tasks and instruction that meet the criteria of being authentic.

Face validity involves making a determination of whether an instrument appears reasonable “on its face.” Typically face validity is determined by experts familiar with

the constructs being measured. If their widespread use is any indication, the AIW rubrics meet the approval of a diverse group of experts. The rubrics have been used as professional development tools for teachers in school systems in Minnesota (Avery, Kouneski, & Odendahl, 2001; Avery & Palmer, 1999). The Gates Foundation used the rubrics to evaluate the performance of reforming high schools (AIR/SRI, 2006, 2007). The Iowa State Department of Education, Michigan State Department of Education, states in Australia (Queensland, New South Wales), and schools in Singapore have adopted the AIW standards (or similar standards) and utilize versions of the rubrics. This suggests considerable face validity. I reinforce the “reasonableness” factor of the rubrics by providing examples of tasks and lessons that scored at different levels on the scales contained within the rubrics in the next chapter. These examples supplement a wide range of examples already available through the various studies I have cited. This documentation enables the reader to judge whether the rubrics are being applied in a valid manner.

Content validity is maximized when a researcher ensures that all the relevant content domains that are incorporated as part of a construct are clearly defined (Trochim, 2006). Strong content validity can mitigate some of the subjectivity associated with rubrics. When applied to authentic intellectual work, content validity means clearly defining what is meant by such domains as higher order thinking and substantive communication. Newmann’s research since the 1960s goes a long way towards meeting this requirement. A brief review of some of the critical works includes research on higher order thinking (Newmann, 1991a), substantive communication (Nystrand &

Gamoran, 1990), and student engagement (Newmann, Wehlage, & Lamborn, 1992). In applying the rubrics, I referred to these studies to clarify points of confusion.

When Newmann field tested these rubrics he consulted subject matter experts (in writing and math for example) to enhance content validity. I ensured content validity in a similar manner. This study served as the pilot for a larger social studies inquiry project involving research sites across the nation. As part of this larger project, I worked with experienced social studies researchers and historians to “norm” the use of the rubrics and establish how to legitimately score social studies instruction based on the degrees of higher order thinking, depth, conversation, and other elements represented in authentic intellectual work. The goal of this process was to ground scoring interpretations in the disciplinary knowledge of history and social studies.

Another way to determine the validity of a construct is to determine if its presence leads to likely outcomes. In the case of authentic pedagogy, Newmann and others hypothesized that higher levels of authentic pedagogy would result in high quality student work as measured by the student work AIW rubric. Several studies have confirmed this relationship (Avery, 1999; King, Schroeder, & Chawszczewski, 2001; Newmann, Lopez, & Bryk, 1998; Newmann, Marks, & Gamoran, 1996).

Given the strong validity of the rubrics, my main task in this study was to ensure that the use of the rubrics conformed to their use in earlier research. My affiliation with the larger national study (Social Studies Inquiry Research Collaborative – SSIRC) focused on the same basic research questions enabled me to attend an authentic pedagogy workshop by Dr. Bruce King. Dr. King is one of the original designers of the AIW rubrics. At this workshop, he shared his knowledge of how to score tasks and

observations. I feel confident that my use of the rubrics reflects the most current thinking on how to measure authentic pedagogy.

**Reliability.** Another important issue to consider when using the rubrics is their reliability. In an effort to enhance the reliability of their use in this study, 22% of the lessons I observed were also rated by my advisor, John Saye. Dr. Saye served as the project director for the SSIRC. The lessons that he observed with me are listed in Table 9. As the table indicates, five of the twenty-three lessons were observed by a second rater. A slightly higher percentage of tasks were also evaluated by a second researcher from the SSIRC project.

The degree of inter-rater reliability for the observations and tasks is depicted in the following ways: 1) by the extent to which scorers had exact agreement on each of the standards and 2) the extent to which agreement was off by 1 point. In every instance, the raters were able to achieve agreement after discussion. Prior research has established a standard of greater than 65% exact agreement and agreement within 1 point to exceed 90% (Newmann & Associates, 1996). Table 10 shows the degree of inter-rater reliability for this study. The lower degree of inter-rater agreement for the substantive conversation and deep knowledge standards was often due to the raters intentionally sampling different groups during group activities. This made it easier to accurately reconstruct classroom events in field notes. However, it also reduced inter-rater reliability when one rater witnessed an interaction the other missed.

Table 9

*Summary of Inter-Rater Reliability Observations*

Date	Lesson
Apr. 2, 2008	Industrial Revolution
Apr. 28, 2008	Reformers lesson
Sept. 22, 2008	An Absolute Monarchy of your Own
Sept. 29, 2008	Declaration of Independence Activity
Classroom Video	British Imperialism in India

Table 10

*Inter-Rater Agreement on Instruction and Assessment Tasks*

	Exact Agreement (%)	Exact or Off by 1 (%)
<b>Instruction (N= 5 lessons/22% of total)</b>		
Standard 1: Higher Order Thinking	80	100
Standard 2: Deep Knowledge	60	100
Standard 3: Substantive Conversation	40	100
Standard 4: Connectedness to the Real World	80	100
<b>Tasks (N= 6 tasks/25% of total)</b>		
Standard 1: Construction of Knowledge	50	100
Standard 2: Elaborated Communication	100	100
Standard 3: Connection to Students' Lives	100	100

**Characteristics of the Task Rubric.** The structure of the task rubric can be seen in Appendix E. The task rubric used in earlier AIW studies included seven standards organized into three broad categories (Newmann & Associates, 1996; Newmann, Secada, & Wehlage, 1995). Later studies revised the rubrics (Newmann, Bryk, & Nagaoka, 2001, see note 10) based on input from experts in specific disciplines. The scoring rubric used in this study included three standards: Construction of Knowledge, Elaborated Communication, and Connection to Students' Lives. Each standard has three levels with the exception of elaborated communication which has four. Tasks which score high in

the construction of knowledge category require students to “interpret, analyze, synthesize, or evaluate information, rather than merely to reproduce information” (Newmann, King, & Carmichael, 2007; Schroeder, Braden, & King, 2001, p. 31). This might involve defending a position on a particular issue or developing a solution to a problem. The elaborated communication standard is a measure of the extent to which students must explain their understanding of the social studies concepts embedded in any particular task. This standard can be met in a variety of ways to include writing, oral presentations, and projects. The final category on the task rubric measures the extent to which the assignment has a connection to students’ lives. High scoring tasks on this standard must do two things: engage students in a problem or issue that has relevance in the real world and provide students with an opportunity to relate to it personally. All of these standards are important and a task only achieves high levels of authenticity by scoring well in each area.

**Characteristics of the Instruction Rubric.** The instruction rubric includes four standards: Higher Order Thinking Processes (HOTS), Deep Knowledge, Substantive Communication, and Connectedness to the Real World. Each of the scales on the rubric has five levels. As in the task rubric, lessons need to score well on each standard to achieve a high level of authenticity. The first standard, higher order thinking, is demonstrated when students are asked to actively manipulate information to solve problems that cannot be solved by simply recalling previously learned material (Newmann, 1991a; Newmann, King, & Carmichael, 2007). This involves engaging students in such processes as analysis, synthesis, and evaluation. The second standard is deep knowledge. A lesson features deep knowledge when sustained attention is given to



a significant disciplinary topic and students are able to demonstrate a thorough and complex understanding of the problem or topic under consideration. The substantive conversation dimension of the AIW framework is a scale that “measures the extent of talking to learn and to understand in the classroom” (RISER, 2000, p. 6). In evaluating this standard, I looked for discussions that featured sustained dialogue focused on disciplinary topics and concepts. Ideally, the dialogue included higher order thinking, the sharing of ideas among participants, and development of coherent understandings. The final standard is connectedness to the real world. In order to score high in this category, teachers must successfully establish the relevance of a lesson to life outside of school. In addition, students must show a personal interest in the topic and attempt to use their knowledge to influence a larger audience other than their classmates (Newmann, King, & Carmichael, 2007).

**Applying & Scoring the Rubrics.** In applying the rubrics in this study, I first asked teachers to submit three tasks that they believed demonstrated their students thinking at a high level about the subject matter of their course (see Appendix G). I referred to these tasks as their most “challenging” in conversations with teachers when asked for clarification about what to submit. I established “curricular validity” by collecting tasks that were designed and/or used by the social studies teachers at the study schools (Ladwig, Smith, Gore, Amosa, & Griffiths, 2007, p. 4). The tasks could be created by someone else (i.e. History Alive) or even be the same as another teacher as long as they represented the teacher’s perception of an assignment that required students to demonstrate thinking at a high level.

Once tasks were submitted, I interviewed or emailed teachers to gain a better understanding of the broader context of how the tasks were used as part of instruction. The intent was to try to connect the three observations directly to the task or to the instruction immediately preceding the tasks. Another goal was to try to set up observations that spanned the course of a semester to provide teachers with a better opportunity to demonstrate the standards associated with authentic pedagogy. If a teacher taught both advanced placement courses and general level courses, I tried to observe at least one class of each type. If a teacher had three general level classes and one advanced placement course, then my observations were weighted more heavily towards the general level classes. Finally, I also took into consideration that a teacher might teach the same lesson differently to different class periods or blocks. In order to address this concern, I attempted to observe different blocks for each teacher. I also asked teachers about this issue during the interview.

Most of the important guidelines for scoring the tasks and instruction are provided on the rubrics themselves. However, a few points should be emphasized. First of all, scoring proceeds from the bottom category and moves up the scale for each standard. In scoring a task or an observation, the next level in a category is assigned only when sufficient evidence is provided to indicate that all of the requirements for the next level are met. When in doubt, the procedure is to score down. Tasks are scored based on the materials provided by the teacher. The interview data and observations yielded additional insights into the dominant expectations a teacher had for any particular task. It was for this reason that I tried to score the tasks after the observations were complete.

The instruction score is based entirely on what is observed during the course of a single class period.

The process for scoring instruction is fairly complicated when recording equipment or multiple observers are not available. My field notes usually included as much dialogue as I was able to capture, my comments or thoughts during the lecture, a class diagram with symbols to represent each student, and a marking system to try to record patterns of conversation. Once I observed a lesson, I sat down as soon as possible to complete my field notes while the information was fresh on my mind. I would go through my notes and highlight or mark areas that represented higher order thinking, students demonstrating depth of knowledge, areas of substantive conversation, and any attempts by the teacher to connect the lesson to the real world. The final step was to assign scores for each standard along with a written justification for each scoring decision.

The mathematical process for scoring requires the development of a composite authentic pedagogy score for each teacher. The scoring of teacher tasks is relatively easy. The task rubric is broken down into three components: Construction of Knowledge, Elaborated Communication, and Connection to Student's Lives. Each category is based on a three point scale except for elaborated communication which extends to a 4. The scores on the three criteria are added together to achieve the authentic pedagogy score for a particular task. Possible scores therefore range from 3 to 10. The scores on each of the three most challenging tasks are averaged to obtain the overall score which is carried forward to the equation used to calculate the final authentic pedagogy score.

The observation rubric is a little different from the task rubric. It has four components: Higher Order Thinking, Depth of Knowledge, Substantive Conversation, and Connectedness to the Real World. Each scale has five levels. Scores for each category are added together to form the overall score for each observation. Scores range from 4 to 20. Once the scores on the three observations are determined, they are averaged to obtain the overall observation score.

The final authentic pedagogy score is calculated by adding the average observation score with the average task score. Once this score is determined, a tenth grader's scores on the designated achievement measures can be compared with the intellectual rigor of the pedagogy he/she experienced in social studies during the ninth and tenth grades (Newmann, Marks, & Gamoran, 1996, p. 16). Additional sub-analyses were conducted on the final authentic pedagogy score for each teacher to determine whether task scores or instruction scores had a greater impact on the dependent variables.

**Determining student prior knowledge.** Data from four different sources were collected to control for student prior knowledge and abilities that have the potential to influence outcomes on the Alabama High School Graduation Exam and the higher order essay assessment. These measures included student end of semester grades in social studies for their eighth, ninth, and tenth grade years. It also included several reading achievement measures because Newmann and others believe that strong readers have an advantage on standardized tests regardless of the content area being assessed (Newmann, Bryk, & Nagaoka, 2001). The first reading prior achievement measure was derived from the Alabama Reading and Mathematics Test (ARMT). This test is administered in the

eighth grade and provides scaled scores that range from level 1 to a level 4. The second reading prior achievement measure was the Stanford 10. It was also administered in the eighth grade. Finally, the Alabama High School Graduation Exam includes a reading component that students take in the tenth grade during the same week that they take the social studies exam. Ultimately, due to the high number of predictor variables and small teacher sample, I only incorporated prior grades into my statistical analyses. This measure was the best determinant of student prior knowledge in social studies.

**Assessing Student Performance.** Two instruments were used to measure student achievement. The first was the Alabama High School Graduation Exam (AHSGE) which best captures student retention of lower order factual content knowledge and basic social studies skills. The second instrument was a researcher developed editorial writing assessment used to measure higher order thinking objectives. The graduation exam was an appropriate instrument for this study because it is the high-stakes social studies test that all high school students must take in the state of Alabama. The test is based on the Alabama Course of Study for Social Studies (Morton, 2004). According to a personal email communication from Dr. Gloria Turner who served as the Director of Assessment for the Alabama State Department of Education, the content standards are “considered to be minimum, required, fundamental, and specific” (G. Turner, personal communication, February 11, 2008). No actual versions of the test have been released to the public making it difficult to conduct a good content analysis. The state has released eighty-four sample item specifications that let students know the general format of the test as well as

the relative weight given to each objective in the social studies curriculum (Richardson, 2000).

I used the item specifications bulletin to analyze the objectives, questioning format, and eligible content to see how they relate to the authentic intellectual work criteria. This analysis, described in chapter 5, and Dr. Turner's correspondence make me confident in describing the test as a measure of lower order content knowledge. Dr. Tommy Bice, the Assistant Superintendent of Education in Alabama also confirmed this in a speech he gave at the Alabama Social Studies Conference in October, 2008. It is therefore the most appropriate instrument to use for research question two.

The test itself covers seven U.S. history standards encompassing America's exploration to World War II. The 10<sup>th</sup> grade curriculum in Alabama only consists of U.S. history through 1877 and therefore does not cover all of the material on the test. Students would have last experienced post-1900 U.S. history content during the sixth grade. The test consists of 100 multiple choice questions each worth 1 point. The results, once scaled, range from 200 to 800 points. The mean score on the test is 500 with a standard deviation of 100 (S. Dubose, personal communication, 2006). Information on the reliability and validity of the test is not readily available to the public.

The high school graduation exam for tenth graders is meant to be a practice run to allow students to become familiar with the test. However, schools obviously want as many students as possible to pass to eliminate the "train wreck" effect that can happen in the later grades when many students still need to pass the exam. The main test administration takes place during an entire week of the Spring semester. Students typically take a graduation exam each day with the testing period lasting all morning.

Students usually take the tests in an assigned classroom accompanied by at least two test proctors. The state has strict testing procedures in place to prevent cheating and encourage standardization in how the test is administered. The week offers students a strong break from the routine. This can impact student motivation, especially when the social studies exam is given later in the week. Also, some tenth graders may lack the sense of urgency felt by the seniors to put forth their best effort.

The higher order thinking assessment designed for this study provided an additional measure of student learning with a focus on several goals that are largely omitted from the graduation exam. The two higher order instruments (one U.S. History, one A.P. European History) were meant to determine the extent to which students were able to analyze arguments made in source documents, weigh competing arguments to arrive at a decision, use historical evidence and prior knowledge to construct a persuasive argument, and apply historical knowledge and critical reasoning to contemporary issues.

The first instrument that I developed was administered to the regular 10<sup>th</sup> grade U.S. history classes. I asked the social studies teachers to select a common topic for the exam; preferably one later in the semester to maximize the potential benefits of instruction on students' performance. The teachers chose Manifest Destiny. This unit was the final unit in the semester before exam review for many of the students. I decided to have students consider the concept of Manifest Destiny through an analysis of the Mexican-American War. In designing this instrument, I kept several things in mind. I did not want to penalize students or restrict their ability to demonstrate higher order reasoning simply because their teacher didn't spend as much time on the Mexican-American war. As a result, I provided students with two resources to assist their

thinking: a timeline of critical events associated with the war and excerpts from two primary source documents. Students with greater prior knowledge could probably do more with these resources. However, I anticipated that a student with a basic understanding of Manifest Destiny, accustomed to classroom experiences that required critical analysis and higher level thinking, would be able to use the documents to comprehend the potential implications of this ideology and frame an argument that scored well.

The Manifest Destiny instrument is included in Appendices H and I. The instrument included two parts. Part I was a structured essay editorial where the student assumed the role of a journalist from the 1840s. The central question asked: *Is using Manifest Destiny to justify war [in Mexico] a violation of American ideals or does pursuing Manifest Destiny in Mexico ultimately promote the greater good?* Students answered this question while adhering to a format that required them to not only lay out their position, but to also address opposing points of view. Part II of the assessment was where students applied their knowledge of Manifest Destiny to contemporary times. The question asked the following: *Consider the role of the United States in world affairs today. Does America still have a special destiny or mission in the world? If so, what is it and how should it be accomplished? If not, explain why you think it does not.* This part of the assessment was used as way to examine the connections students were able to make between an historical topic of study and contemporary times.

This was a valid assessment of student learning for several reasons. First of all, the content adhered to the required tenth grade curriculum which covers U.S. History through 1877. Since the topic was suggested by the social studies teachers, I know that



the students received instruction pertaining to Manifest Destiny and, to at least some extent, the Mexican-American War. The instrument was created in conjunction with my advisor and reviewed for face validity by two other social studies teacher educators and a secondary social studies classroom teacher. Each of these social studies professionals has significant experience and expertise in the field. In their opinion, the instrument included appropriate content that was realistically formatted for students at this grade level. They also found the instrument to be an adequate measure of the types of higher order thinking that I envisioned.

Having established content and face validity, the next concern was whether this instrument truly measured the types of higher order thinking processes commonly associated with authentic tasks. The instrument evaluated some of the same lower order knowledge as the graduation exam (i.e. the definition of Manifest Destiny). However, it provided a greater overall challenge by requiring students to take a position on a central question using extended writing. In order to do this effectively, students had to be able to extract important details from the supporting materials (timeline, documents) and synthesize them into a coherent argument. This required not only understanding the viewpoints represented in the documents, but also connecting the information to prior knowledge. Since students were ultimately evaluating the justness of America's policies, their answer required logical reasoning, generalizing from evidence, making distinctions, and a host of other possible higher order processes. In recognizing opposing arguments and responding to them, students were also demonstrating their ability to use dialectical reasoning. Dialectical reasoning is a central component in the process of building a decision-making model for critically evaluating public issues. It simply involves being

able to critically analyze a problem and understand perspectives different from your own (Parker, 1989, p. 9). Finally, by adding a real world component where students connected the principles of Manifest Destiny to modern times, the essay included all three elements of an authentic task (construction of knowledge, elaborated communication, and a connection to students' lives). The task was challenging for tenth graders and adults alike.

In order to check the reliability of scoring on the higher order tasks, I had another doctoral student in social science education evaluate a random sample of editorial using the same rubrics. The percentage of agreement with my original scores was 55% on the German Unification editorial. Table 11 provides a breakdown of the inter-rater agreement for the various rubric categories.

The degree of inter-rater reliability on the AP editorials for the persuasiveness standard (Part I – 26%) is a bit misleading. In certain cases, the scores that I assigned disagreed with those of the other rater, but they both represented “minimal” persuasiveness (1 vs. 2 on the rubric). If you count minimal scores, whether 1 or 2, as agreeing, then the level of exact agreement rises to 52%. Furthermore, out of the 23 editorials examined, we agreed that the majority (87%) represented adequate persuasiveness at best. The position statement IRR score (43%) is also quite low. In our follow-up conference I determined that the other rater had misinterpreted the standard and was counting any statement that argued for unification as providing a clear position on the question. I was looking for students to argue for a particular vision of unification (i.e. small German solution, Germany with Austria, etc.). This misunderstanding definitely caused our level of agreement on this standard to be artificially low.

Table 11

*Inter-Rater Agreement on Higher Order Editorial Tasks*

	Exact Agreement (%)	Exact or Off by 1 (%)
<b>German Unification AP Task (N=23; 25% of total)</b>		
<b>Part I</b>		
Standard 1: Position	43%	N/A
Standard 2: Historical Context	65%	97%
Standard 3: Persuasiveness	26%	78%
Standard 4: Low-Level Dialectical Reasoning	30%	83%
Standard 5: Quality of Final Position	70%	100%
<b>Part II:</b>		
Standard 1: Decision-making	87%	N/A
Standard 2: Persuasiveness	65%	96%

The second testing instrument, used for the Advanced Placement European History classes, focused on German unification. It is included in Appendices J and K. This instrument adheres to the same basic format as the U.S. History assessment. The topic of German unification is routinely covered in the AP curriculum and was suggested by the AP teachers involved in the study. The instrument was created in conjunction with my advisor and reviewed for face validity and content validity by a social studies teacher educator, a doctoral student with experience teaching a similar course, and a secondary social studies classroom teacher at the study school. In their opinion, the instrument included appropriate content and was a realistic assessment for the target population of students. They also felt it measured the higher order thinking objectives for which it was designed.

The essay question for AP students was the following: *Should the unification of all Germanic peoples within one nation be endorsed (supported) by the German people? Would other nations likely support it?* Students were provided a timeline of significant

events leading up to 1870; the decision point in this exercise. They were also provided with primary documents which advocated unification on different terms. These were used to evaluate potential courses of action (i.e. Lesser or Greater Germany?) for solving the German question. Students essentially had to take a stand on the principles that should guide unification. Should unification be based on nationalism or self-determination of peoples? Of course, students could also argue against unification.

The AP essay also included a connection to contemporary issues. In this case, students were asked to answer the following question: *To what extent, if any, should the U.S. support the ambitions of ethnic, cultural, or religious groups seeking to secure their own nation-states today?* The students were provided with several examples of groups seeking independence to help them better understand the issue and frame a response (i.e. efforts to secure a Palestinian state; Kurdish uprisings in Iraq, etc.). This task places similar cognitive demands on students as the previously described U.S. history assessment. Students must construct persuasive arguments regarding German unification and modern nation building. In doing so, they engage in extended writing about an issue with real world significance. The task can therefore be considered authentic.

Both higher order assessments were administered by the classroom teachers that participated in this study. I provided instructions for the teachers to read to their students as part of this process. These instructions are provided in Appendix L. Students had one hour to complete the assessment. Once they finished the assessment, the exams were collected by the classroom teacher and forwarded to the department head. I then picked up the exams for scoring. The editorials were anonymous to me since they contained only a student number and no reference to the class or the teacher. I scored all of the

editorials before entering the results in my database. This helped to ensure my scoring was not biased to favor students from a particular teacher.

The rubric that I developed for the editorials was most influenced by the scoring guide created by Newmann to evaluate persuasive writing (Newmann, 1990). I also decided to incorporate scoring elements from two other relevant studies that measured competencies such as dialectical reasoning that were present in the editorial task (Parker, Mueller, & Wendling, 1989; Saye & Brush, 1999b). The AP editorial rubric evaluated students in five categories for Part I (See Appendix M). The first category was the position statement. Students received a point if they provided a clear statement of one to two consecutive sentences that explained their stance on the question. For instance, “The unification of all German peoples within one nation should be endorsed by the German people because....” In many cases, I was able to infer a student’s stance based on statements made throughout the editorial. However, I asked students to provide an explicit statement. If students did not follow the instructions, they did not receive the point for this category.

The next scoring category evaluated how well students set up their editorial in the first paragraph. The historical context scale extended from 0 to 2. Students who provided no background information whatsoever received a 0. Level one scores required at least some historical context. This typically consisted of one or two sentences of background information that closely followed the language used in the timeline provided with the task. Simply mentioning relevant events like the unification wars or Bismarck’s influence on the unification process would qualify. Level two scores were reserved for students who demonstrated some knowledge beyond what was provided on the timeline

or for students who provided a more detailed introduction that used language that differed from the timeline. Level two introductory paragraphs had to clearly present accurate information to set up the student's position statement.

The next category on the scoring rubric was persuasiveness. The persuasiveness score was derived from a close reading of the entire editorial even though paragraph two was the main paragraph designated for supporting arguments in the assignment instructions. In order to evaluate persuasiveness, I generated a list of plausible arguments that could be made to support either side of the focus question. This list was consulted when making decisions about the number of distinct arguments being made in any particular editorial.

The persuasiveness scale had five possible levels. Editorials that scored at the first two levels were considered minimally persuasive and unlikely to persuade the reader. In order to receive a "1", the student had to provide one persuasive argument to back up his/her stance on the question. The argument could conceivably consist of only one sentence and did not require any elaboration or inclusion of historical evidence from the source documents. Level "2" scores were usually assigned to students who misunderstood the question. These students provided multiple arguments related to the pros or cons of unification instead of defending a position regarding nationalism and the territory a unified Germany should include. Students who made this mistake could not earn a higher score than a 2.

A level 3 "adequate" persuasiveness score was assigned when students were able to provide two reasons to support their position or one reason that included useful elaboration. The main consideration in assigning a 3 was whether the editorial "had a

chance of persuading the reader” given the elaboration provided by the student. The next scoring level of “elaborated” required the student to provide either more elaboration (i.e. citing historical evidence, use of examples, etc.) or additional reasons to back up his/her stance (at least 3). Elaborated editorials were considered likely to persuade the reader. The level 5 scoring category was referred to as “exemplary”. Level 5 scores were even more persuasive than the level 4 editorials mainly due to especially clear and coherent argumentation. These editorials were polished enough (i.e. no major grammatical mistakes) to be considered for public display as outstanding accomplishments for tenth grade students. In general, the persuasiveness score reached at least the adequate level when students are able to accurately reference the primary source documents or integrate valid historical analogies or examples into their writing. However, students could also hurt their score by providing inaccurate statements or statements that undermined their overall argument.

The final two scoring categories were low-level dialectical reasoning and quality of the final position: a standard that measured the ability of students to engage in high level dialectical reasoning while crafting a persuasive closing argument. To engage in low level dialectical reasoning, students had to correctly identify and explain opposing viewpoints. The scale for this category went from 0 to 3. The lowest score of “1” required students to correctly state an opposing view in minimal terms. For example, a student who argued for the “Greater German solution” might discuss the opposing view of how nationalism could promote further warfare as Germany sought to incorporate German-speaking territories not presently under their control. I considered a response “minimal” when the student provided a single sentence explanation of the opposing

perspective. Students who provided multiple opposing viewpoints that were briefly articulated received a two. The highest score in this category was reserved for students who explained at least one opposing view in greater detail by providing examples and evidence from the supporting documents.

In looking at the quality of the student's final position, I analyzed the level of persuasiveness and dialectical reasoning demonstrated in the fourth paragraph. I was looking for students to frame their closing arguments around a thoughtful response to the critics. Students also needed to restate their thesis and most significant points. Students who did not provide a fourth paragraph received a 0 for this scoring category. A "1" score required students to respond to the arguments of critics and briefly mention or restate at least one key point from the editorial. This "adequate" conclusion represented a minimal response that really didn't add anything to the persuasiveness of the overall editorial. A "2" conclusion required either a particularly strong (as in elaborated and persuasive) response to the critics or a more detailed summary of the key arguments from the editorial. A level 2 paragraph was more persuasive than a level 1, but did not feature the advanced dialectical reasoning needed for the highest score in this category.

Level 3 scores were reserved for conclusions with tight argumentation and genuine consideration of opposing views. After reading a paragraph at this level, the reader should have very few, if any, unanswered questions. Advanced dialectical reasoning is demonstrated when students fairly characterize the view of critics and respond to them in a thoughtful and respectful manner (i.e. While my opponents make some valid points, I still feel that....).



Part I of the Manifest Destiny editorial was evaluated using a similar rubric (see Appendix N). The range of scores in each category was identical. In the historical context category, I looked for students to provide at least some information about the border dispute that directly preceded the Mexican-American War to receive the maximum points. A level 2 score on the persuasiveness scale once again mainly captured those students that didn't quite understand the question. In this case they didn't provide any comments related to Manifest Destiny, choosing instead to provide multiple arguments for or against going to war with Mexico. In order to receive a higher persuasiveness score, students had to relate their response to the concept of Manifest Destiny. The wording in the "quality of final position" category is slightly different for this rubric mainly because many of the students responded to critics in paragraph 3 instead of the final paragraph. However, students were still evaluated on persuasiveness and advanced dialectical reasoning.

In Part II of both editorials, the goal was for students to connect their historical knowledge to a modern issue. I evaluated the question provided to the AP students using two scoring categories: decision-making and persuasiveness. The decision-making category was similar to the position category in part I. I looked for how clearly the student defined his/her position on the question of whether the U.S. should support the formation of new nation-states. Students who took a clear stance received a point. The persuasiveness scale was essentially the same as the one used in part I, but had four levels instead of five (the "2" score was removed from the Part I scale).

The question for the regular students was evaluated based on the extent to which students' appeared to make connections in their response between modern ideas of

American exceptionalism and “mission” and the historic concept of Manifest Destiny. Scores were broken down into three levels: 0 = no connection, 1= possible connection, and 2 = explicit connection. The “no connection” responses didn’t provide any indication of whether the student recognized any parallels between U.S. actions today and the ideas associated with Manifest Destiny. These responses were sometimes completely off topic reflecting a misunderstanding of the question. The “possible connection” score was assigned to students who made some valid historical connections in their response or perhaps touched on some of themes associated with American exceptionalism. The explicit connection score was reserved for students who compared America’s modern mission (as they perceived it) directly with its historic destiny as it was conceived by advocates of Manifest Destiny in the 1800s. Students who referenced Manifest Destiny in their response in some valid way could receive a “2” score.

***Researcher as Instrument.*** As in any study that includes a qualitative component, the researcher is an important instrument to consider in the analysis. Most of my professional background in teaching social studies is based on an inquiry model. I have a bias towards this instructional approach and feel that its objectives are largely at odds with the type of learning encouraged by standardized, multiple-choice tests. Recognizing this bias, I attempted to mitigate the impact of my personal feelings by confirming my analyses with other researchers (inter-rater reliability) and through the use of multiple sources of data (tasks, interviews, observations, standardized tests) which support triangulation to corroborate findings made in the study.

## **Study Phases**

This section describes the various phases of the study along with the data collection process. The study was broken down into four phases. The first phase was the planning and design refinement stage of the research. This involved selecting a meaningful topic, clearly defining the purpose of the research, and developing the research questions. During this phase I also prepared the basic study design while obtaining approval to proceed with the study from the school system and the Institutional Review Board (IRB). In Phase II, I implemented the study by beginning the process of collecting teacher data. At the same time, the school system began to organize the necessary student data into spreadsheets for the 2007/08 data set. Phase III began in the Fall of 2008 and continued through the school year. During this time I collected the remaining teacher data while also administering higher order essays. I also collected the student data (class rosters, demographics, test scores) for the 2008/09 data set. The final phase was the data analysis stage. While this is presented as a distinct phase, it actually occurred throughout the study. This information is summarized in the Figure on the next page.

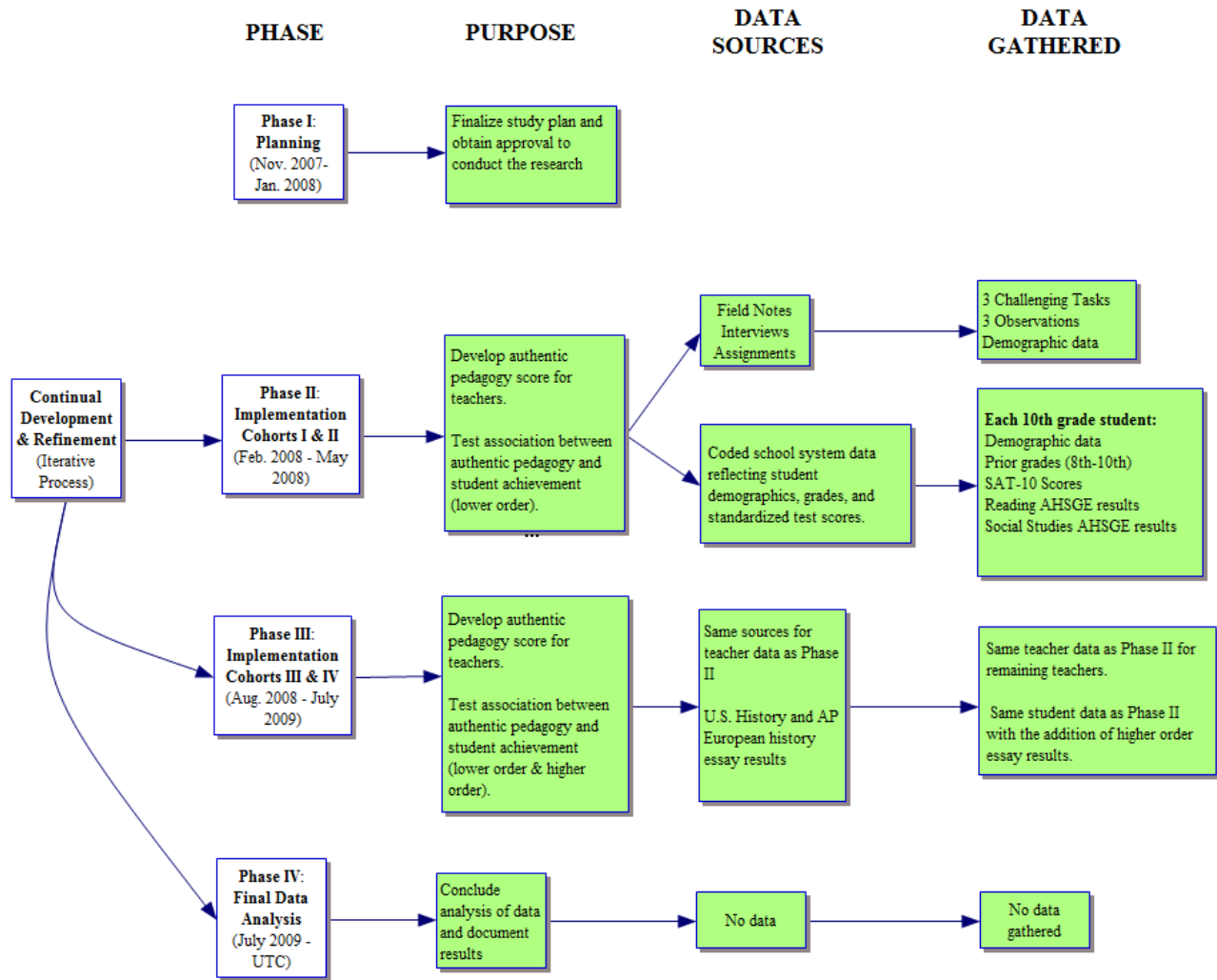


Figure 2. Summary of Research Phases

**Phase I: Planning and Design Refinement.** This study was first conceptualized in 2005. During the process of developing the topic, I did a review of the relevant research and narrowed down a list of potential research questions. The specifics of the study design evolved over time and were finalized in 2007. A four hour consultation with Dr. Bruce King at the November annual conference of the College and University Faculty Assembly (CUFA) of the National Council for the Social Studies (NCSS) helped with this process. During the Fall semester of 2007, I approached school system officials with the idea for the study. I met with the assistant superintendent, central office personnel, and the school principals. Once I had IRB approval and their support, I began the process of recruiting the 9<sup>th</sup> and 10<sup>th</sup> grade social studies faculty.

In January 2008, I conducted separate meetings with the 9<sup>th</sup> grade social studies faculty at the junior high and the 10<sup>th</sup> grade social studies faculty at the high school. I went over the details of the study using a briefing script prepared for the IRB as a guide (Appendix B). Teachers were encouraged to ask questions and informed of their right to not participate or opt out of the study at any time. Each teacher agreed to participate and signed the consent form.

I did not have to recruit students since the student data are anonymous secondary data that do not require participant consent. The student data consisted of demographic and achievement reports already collected by the system or collected as part of a system sponsored pilot assessment. I obtained student results in a coded form that prevented me from knowing any student names. Observations were not videotaped or recorded in any

way other than through general field notes, so student anonymity was maintained throughout the study.

As the study progressed, I had to make occasional adjustments to the initial plan based on unforeseen circumstances (i.e. changes in teachers, etc.). I also worked through the process of finalizing specific instruments or protocols for later stages of the study. The process of design refinement was therefore initiated in the first phase and returned to throughout much of the study.

**Phase II – Implementation – 2007/08 School Year.** Phase II began in February 2008. The first step in the data collection process for this phase involved the collection of three tasks from the study teachers. This was a departure from previous AIW studies in several ways. Earlier studies required teachers to submit examples of typical assignments and challenging assignments. During the AIW workshop, Dr. King noted that this did not substantially alter the types of assignments submitted by the teachers. On his recommendation, I decided to forgo the request for typical tasks in order to have teachers focus on choosing the three tasks that they felt best represented their students thinking at a high level. The language used in the protocol requesting these tasks (see Appendix G) mirrors what was used in the previous authentic intellectual work studies conducted by Newmann. Teachers were told that they could either submit their tasks electronically or arrange a time for me to collect them in person. I provided a deadline for submitting tasks of February 15, 2008. Some teachers needed additional reminders causing tasks to be submitted sporadically throughout the semester.

Another decision made in consultation with Dr. King involved the amount of data to collect from teachers. Newmann and King tried different approaches for collecting teacher data and found that additional tasks and observations (beyond three each) did not enhance their ability to differentiate between teachers. The adoption of the three task/three observation design thus stems from “lessons learned” from earlier AIW research and my desire to minimize the demands on teachers during this study.

One other difference between this study and previous ones of this type was the attempt to link the tasks submitted by teachers to observations. Dr. King recommended this due to difficulties associated with interpreting tasks as stand-alone artifacts. The evaluation of tasks in this study generally followed the observations. I restricted my analysis of tasks primarily to the materials a teacher submitted. However, in judging the overriding instructional intent of a teacher for a particular task, I did take into consideration insights from the lesson observation.

I negotiated with teachers to schedule the lesson observations. Written tasks, such as essays or reports, provide little opportunity for me to observe criteria specified on the AIW observation rubric. In order to afford each teacher the opportunity to score well on the observations I employed the following strategy: when the task involved a debate, simulation, or some other form of “live” presentation, I observed the actual class period associated with the task unless the teacher had a specific rationale for observing another day. However, if a task involved a written assignment that was difficult to observe, I negotiated with the teacher to observe the day that best demonstrated how students were prepared to complete the task.

The collection of some teacher data had to be pushed back to phase III for several reasons. First, three teachers had interns during the Spring 07 semester. The interns were teaching units that in some cases corresponded with the challenging task(s) the teacher wished to submit. It was also evident that scheduling observations was going to be difficult due to the intern observation schedule. The second factor that caused the collection of some teacher data to be postponed related to delays in initiating the study. Data collection did not formally begin until mid-February. Some teachers submitted challenging tasks that they had already taught to their students. I tried to allow teachers to stick with their original selection unless the teacher legitimately felt another task was just as challenging. Also, I wanted to observe tasks/lessons that were spaced throughout the semester realizing the potential increase in difficulty for teachers to score well on the AIW standards early in the semester. Finally, despite repeated communication attempts, some teachers did not provide their tasks in a timely fashion or did not respond to attempts to schedule observations.

In addition to collecting tasks and conducting observations, I also completed interviews with some of the study teachers (see interview script, Appendix A). My study design, approved by the IRB, only permitted one brief interview of approximately fifteen minutes instead of the pre/post interview schedule eventually adopted by the larger SSIRC study. I preferred to conduct the interview after all the teacher data had been collected. However, it was more difficult to negotiate the observation dates with some teachers than I originally anticipated. As a result, I set up some meetings during phase II when observation dates were finalized and the interview questions were completed simultaneously. I also went ahead and did interviews with some teachers after the



majority of their data had been collected. I was unable to conduct an interview with one teacher who retired during the study. I was also only able to collect background data from another teacher.

Finally, I worked to improve the inter-rater reliability of the rubrics during this phase and subsequent phases. A steering committee conference of the Social Studies Inquiry Research Collaborative was held at Auburn University on March 5-7, 2008. This meeting included substantial time for practice scoring with the AIW rubrics. The observation rubric was used in conjunction with video footage and an actual classroom visit. A subsequent meeting, which I was not able to attend, was held at the American Educational Research Association (AERA) conference on March 26. This meeting also included scoring practice with an on-site evaluation of a class in New York. I benefited from the minutes and discussion that came out of this conference. Following each of the sessions noted in Table 12, I revisited my field notes for completed observations. My field notes included a specific rationale for each scoring decision on the AIW rubric. When my thinking was “normed” by more precise interpretations of the rubrics, I could easily determine whether a score needed to be adjusted.

**Phase III – Implementation – 2008/09 School Year.** During phase III, I focused on collecting the remaining teacher data, analyzing phase II data, and creating/implementing higher order assessments. I worked on the Manifest Destiny assessment in the Fall and provided it to the tenth grade U.S. history teachers in December. A total of 184 students took the exam. The exam for the Advanced Placement students was administered during the Spring semester.

Table 12

*Summary of Inter-Rater Reliability Sessions*

Data	Location	Purpose	Role
Nov. 2007	San Diego, CA	CUFA meeting at NCSS – Consultation with Dr. Bruce King.	Participant
Mar. 5-7, 2008	Auburn, AL	Steering Committee of SSIRC – Task and observation rubric practice using actual tasks, video, and an on-site classroom visit.	Participant
Mar. 26, 2008	New York	AERA Conference – Task and observation rubric norming using actual tasks, video, and an on-site classroom visit.	Access to minutes
June 19, 2008	Internet Video Conference	Steering Committee of SSIRC – Observation rubric norming using Geovanis video.	Access to minutes
July 24, 2008	Internet Video Conference	Steering Committee of SSIRC – Observation rubric norming using Eubanks video.	Participant
Sept. 22, 2008	Auburn, AL	Norming session of task and observation rubrics.	Participant
Oct 3, 2008	Auburn, AL	Alabama Conference of SSIRC – Observation and task rubric norming	Participant
Nov. 13-14	Houston, TX	CUFA meeting at NCSS – Task and Observation rubric norming	Participant
Dec. 15, 2008	Internet Video Conference	Online Norming session	Participant
Jan. 17, 2009	Charlottesville, VA	CUFA Retreat at the University of Virginia	Access to audio recording
Apr. 15, 2009	San Diego, CA	AERA Conference – Task rubric norming using a task collected by a SSIRC researcher.	Access to minutes

**Phase IV – Final Data Analysis.** Data analysis was an iterative process during this study. It began during phase II with the collection of teacher data and continued until the end of the study. Phase IV began in July 2009 with the culmination of the second year of data collection.

## **Data Analysis Procedures**

The data analysis process involved the analysis of student and teacher data to ascertain the impact of instruction on social studies learning outcomes. My analysis focused specifically on five research questions. The questions are listed below:

*Research Question 1:* To what extent do teachers utilize authentic pedagogy and how much variation exists within the sample of teachers in this study?

*Research Question 2:* Do students that have been taught by teachers demonstrating higher levels of authentic pedagogy score higher on the AHSGE than students taught by teachers with lower levels of authentic pedagogy?

*Research Question 3:* What is the impact of authentic pedagogy on student performance on an assessment that requires them to apply knowledge from a previous unit to a challenging new task?

*Research Question 4:* Does the ability to apply knowledge in these situations improve with repeated exposure (multiple courses) to classroom experiences that require students to perform challenging intellectual tasks?

*Research Question 5:* To what extent does authentic pedagogy bring different achievement benefits to students of different social and academic backgrounds?

Table 13 depicts the hypotheses associated with each of these questions. It also provides an overview of the data analysis methods.

Table 13

*Summary of Research Questions and Data Analysis Methodology*

Research Question	Hypothesis	Method of Analysis
To what extent do teachers utilize authentic pedagogy and how much variation exists within the sample of teachers in this study?	The mean score of the teacher sample will not reach the mean of the authentic pedagogy scale.	Application of Newmann's task and instruction AIW rubrics. Analysis of descriptive data.
Do students that have been taught by teachers demonstrating higher levels of authentic pedagogy score higher on the AHSGE than students taught by teachers with lower levels of authentic pedagogy?	Students who experience higher levels of authentic instruction in their tenth grade social studies courses will on average achieve higher scores on the graduation exam.	Multiple regression using students as the unit of analysis & a comparison of classes using one-way ANOVA
What is the impact of authentic pedagogy on student performance on an assessment that requires them to apply knowledge from a previous unit to a challenging new task?	Students who experience higher levels of authentic pedagogy in their social studies class will on average achieve higher scores on the higher order essay assessments.	Factorial MANOVA analyzing performance of students who received minimal, limited, or moderate levels of authentic pedagogy

Table 13

*Summary of Research Questions and Data Analysis Methodology (Cont.)*

Research Question	Hypothesis	Method of Analysis
Does the ability to apply knowledge in these situations improve with repeated exposure (multiple courses) to classroom experiences that require students to perform challenging intellectual tasks?	Achievement benefits associated with authentic pedagogy will be enhanced by increased exposure (multiple semesters) to social studies coursework that meets the standards of high quality designated by this construct.	Analysis of post-hoc tests from one-way ANOVA and multiple regression analysis
To what extent does authentic pedagogy bring different achievement benefits to students of different social and academic backgrounds?	Achievement gains associated with higher levels of authentic pedagogy will be equitably distributed among the student population associated with this study.	Analysis of bivariate correlations

**Data Preparation.** Data was received from the school system in the form of spreadsheets that had to be reorganized and merged into a coherent database. The data collection process proceeded incrementally since the study overlapped two school years. Standardized test results were only available at certain times based on the reporting cycle followed by the state. I also had to follow the district schedule to obtain student grades, course schedules, and other information. A consequence of obtaining student data piecemeal was an increased likelihood that the spreadsheets would not match perfectly and therefore some data was missing. Whenever possible, I tried to reconcile discrepancies and obtain missing data from the district. However, the final dataset was still incomplete in some areas. When running statistical tests, I only included students with complete records for the variables under analysis. Once I had a coherent database that incorporated all of the different spreadsheets I had received from the school system, I began the process of preparing the data for analysis in SPSS. I created new categorical variables for race, gender, SES, limited English proficiency, and special education based on the mean Alabama High School Graduation Exam scaled scores of students in these categories.

**Analyzing Teacher Data.** In order to address the first research question, I assigned authentic pedagogy scores to the teachers based on task and observational data. The process for assigning these scores is described in the instrumentation section of this chapter. The final authentic pedagogy scores (average task score plus average instruction) were used as the basis for categorizing teachers into four groups: minimal authentic pedagogy, limited authentic pedagogy, moderate authentic pedagogy, and

substantial authentic pedagogy. These categories were developed by evenly breaking the final authentic pedagogy scale (7-30) into quartiles as follows: Q1 = between 7 and 11.99, Q2= between 12 and 17.99, Q3= between 18 and 23.99, Q4= above 24

**Analyzing Student Learning Outcomes.** Multiple levels of analysis were necessary to analyze the impact of authentic pedagogy on student performance on the social studies graduation exam. This was mainly due to the small size of the teacher sample at grade 10 (N=4) which made it more difficult to determine if the effects of instruction were significant. My initial examination of the data focused on students as the unit of analysis (N= 805) and utilized a multiple regression model whereby student variables known to influence achievement were controlled to reveal the independent effects of authentic pedagogy on the dependent variable (Alabama Graduation Exam scaled score). The independent variables listed in Table 14 were analyzed when addressing research question two and the other research questions.

Table 14

*Overview of Independent Variables Used During Regression Analyses*

Variable	Coded Name	Value
Gender	Sex	M=Male, F=Female
Ethnicity	Race	W=White, A=Asian B=Black H=Hispanic N=Not Reported
SES	Lunch	1=Free, 2=Reduced, 3=Paid
Disability Status	SpEduc	0=No, 1=Yes
English Proficiency	LEP	0=No, 1=Yes
Course Name	CourseName	AP US USAlt. USCo.
Course Type	CourseType	1=Fall, 2=Spring, 3=All Year
Prior Grades (10)	Average10	0-100
Prior Grades (9)	Average9	0-100
Average Task Authenticity	TaskComposite	3-10
Average Instruction Authenticity	InstructionComposite	4-20
Authentic Pedagogy Score	APScore	7-30

*Note.* All variables were not retained in the final analysis. I recoded the first seven variables as criterion-coded variables using the mean scale AHSGE scores for students in each particular category (i.e. the mean scores for males & females, etc.).

In order to investigate the relationship between the many predictor variables, I conducted the regression sequentially. In step/block one, I entered the student demographics. Then, in step/block two, I entered the prior achievement measure of student grades in social studies. Finally, in step/block three I entered the authentic pedagogy variables.



In trying to determine the usefulness of the various predictors, I utilized the following criteria. First, I verified that the regression model itself had a high F (ANOVA) that wasn't likely to occur often by chance. During each stage of analysis, I examined the  $R^2$  of the predictor variables. A high  $R^2$  with a low standard error of estimate was desired. I also looked for the variables that had the highest Betas with a high t-value that was significant (.05 or less). Finally, I looked for variables with a high semi-partial correlation. This was probably the best indicator because it showed how much a variable contributed on its own to predicting the criterion variable. After looking at each of these indicators, I was able to determine the extent to which each variable influenced the achievement outcomes of this study and to rank order the predictor variables in order of their importance.

I tested four critical assumptions associated with this statistical approach (Osborne & Waters, 2002). I checked to see if the variables were normally distributed, a linear relationship existed between the independent and dependent variables, the variables were reliably measured, and if the residuals were normally distributed among the independent variables. Each of these assumptions was met. I also analyzed the correlation matrix and collinearity statistics within the SPSS reports to ensure that the predictor variables were not highly correlated with each other.

The multiple regression analysis related to this research question determined the extent to which a relationship existed at the individual student level between authentic pedagogy and student performance above and beyond any of the other variables. My conclusions must be viewed as extremely tentative since teacher characteristics were not controlled. Also, the variability of instruction when using students as the unit of analysis

was very limited since all of the students were associated with only four high school teachers.

In order to make a stronger case regarding the impact of authentic pedagogy on student performance I also ran some analyses using the classroom as the level of analysis. I used one-way ANOVAs to compare classes from specific authentic pedagogy categories (minimum, limited, moderate). I was very careful to identify classes for these comparisons that had similar students. I did this by generating contingency tables in SPSS using the crosstab command. The crosstab command results in a Pearson Chi-Square test for each variable making it easy to identify statistically significant differences between classes. In addition to matching classes based on student characteristics (demographics, prior achievement), I also attempted to match classes based on some teacher characteristics. Ultimately, this process held the teacher and student characteristics constant, thus focusing the ANOVA on the specific impact of authentic pedagogy on student performance. The combination of one-way ANOVAs and multiple regression analysis enabled me to more effectively address the second research question.

Research question three focused on the ability of students to apply the knowledge, skills, and dispositions gained from instruction on a particular topic to a challenging new task. In this case, the new task was an editorial writing assignment that required a good deal of higher order thinking and elaborated communication. The task was also structured to measure the ability of students to connect historical knowledge to contemporary events and issues. Higher order tasks of this nature fit more readily into the stated goals of authentic intellectual work. Two different essays were prepared for the students. The advanced placement students taking European History completed an

essay focused on German unification. The regular education students in the U.S. History courses were administered an essay on Manifest Destiny during the Mexican-American War. These essays were evaluated using the rubrics in Appendices M and N.

The procedure for analyzing this research question involved several steps. First, I isolated the sample of students who had taken the higher order editorial by filtering out the other students in the database. Then, I created a new teacher grouping variable based on the authentic pedagogy categories I had previously established (1=minimal; 2=limited; 3=moderate). I combined the students who received minimal authentic pedagogy (from Andy and Jason) into one group. The other two groups (limited & moderate) included one teacher in each group. Next, I ran an analysis using SPSS to determine the extent to which significant differences existed between the three groups of students on certain demographic variables (gender, SES, and ethnicity). This process was essentially the same as what I did for research question two. I generated contingency tables for each variable using the crosstab command. The resulting Pearson Chi-Square was used to identify statistically significant differences between the three groups. I also ran a oneway- ANOVA to determine whether significant differences existed between the groups based on their social studies grades from the current year. The goal of each of these steps was to control for as many factors as possible other than authentic pedagogy that could influence student performance on the higher order editorial. Finally, I ran a MANOVA to test the hypothesis that students who experience higher levels of authentic pedagogy achieve higher scores on the higher order assessment. The dependent variables were the rubric categories associated with the higher order assessment. The fixed factor

was the level of authentic pedagogy students experienced (represented by the three teacher groups I had established).

Although the advanced placement and regular editorials were formatted similarly and the rubrics were virtually the same, I decided to analyze the work separately. I didn't feel confident comparing the performance of these students because I couldn't be certain that the challenge they experienced was the same. I did, however, apply the same statistical procedures to each set of data.

Research question four focused on whether there was a performance benefit associated with taking multiple social studies courses that featured higher levels of authentic pedagogy. In attempting to address this question, I did not observe firsthand the type of instruction a group of students received over the course of two years. Instead, I collected teacher data for the entire sample of 8 teachers for one year and performed my analyses (on two years worth of student achievement data) based on the assumption that teachers do not radically alter their instruction between semesters or consecutive school years. For example, I placed one of the teachers in the limited authentic pedagogy category based entirely on observations during the spring semester, 2008. I made the assumption that students who had this same teacher during a different semester also experienced limited amounts of authentic pedagogy.

I created a new "prior moderates" variable for this research question that was a measure of the number of social studies courses each student experienced that were at the moderate authentic pedagogy level. The prior moderate variable had three possible options. A student with a "0" designation did not experience any social studies courses at the moderate authentic pedagogy level in the ninth or tenth grade. A "1" indicated that

the student had at least one course at the moderate level in either the ninth or tenth grades. Finally, a “2” meant that both the ninth grade and tenth grade social studies courses the student took were at the moderate level.

I ran several one-way ANOVAs to analyze this question. In each case, I factored out students who had more than one social studies course during either the ninth or tenth grades. The first one-way ANOVA included all of the students in the sample. The second removed the advanced placement students and just looked at the impact of multiple years of moderate authentic pedagogy on regular education students. The final ANOVA just featured the advanced placement students. In each instance, the ANOVA provided an indication of whether the overall model was significant. I ran several post-hoc tests to see if any significant differences in performance existed between students with 0, 1, or 2 course featuring moderate levels of authentic pedagogy.

In addition to the ANOVAs, I also ran a sequential multiple regression analysis that was identical to the one described for research question 2. However, instead of including the task and instruction authentic pedagogy variables in step 3, I entered the prior moderate variable. This enabled me to have a measure of the unique impact of the prior moderate variable above and beyond the influence of the demographic and prior achievement variables.

Finally, I was interested in determining the achievement effects of authentic pedagogy for specific subgroups of students. Using the demographic and achievement data obtained from the school system, I ran a series of bivariate analyses in SPSS to determine whether the achievement benefits associated with authentic tasks and authentic instruction were equitably distributed among the students at Central High School. I

analyzed the impact of authentic pedagogy based on gender, race, SES, and prior academic achievement in social studies. Each bivariate analysis resulted in a Pearson correlation statistic that served as the main indicator of whether a correlation was significant. I compared the direction of correlation and the level of significance for each subgroup of students to determine whether certain students (i.e. males) were more or less likely to be advantaged by higher levels of authentic pedagogy.

### **Conclusion**

In conclusion, this was a four phase study focused on understanding the relationship of authentic intellectual work to student learning outcomes (both higher and lower order) in social studies. It was a mixed method analysis of the social studies instruction at two study schools in East Alabama. The study isolated the impact of authentic pedagogy on student performance primarily through regression analyses by controlling for a variety of predictor variables that were likely to have at least some impact on achievement. The study ultimately included 805 students, eight teachers, and data collected over the course of two school years.

## CHAPTER FOUR: TEACHER USE OF AUTHENTIC PEDAGOGY

The purpose of this chapter is to present findings related to my first research question: to what extent do teachers utilize authentic pedagogy and how much variation exists within the sample of teachers in this study? This chapter includes raw scores from my analysis of this question as well as descriptive accounts of teacher practice at various levels of the authentic pedagogy continuum. These accounts are intended to help the reader form a more complete understanding of the types of intellectual challenges students experienced in their history courses. As discussed in the previous chapter, the authentic pedagogy scores are based on an analysis of tasks and instruction. Teachers were asked to submit three tasks that best indicate how well students understand their subject at a high level. These tasks were then each linked to a classroom observation. The observation sometimes featured students actually engaged in the associated assignment. In other cases, I observed the instruction that prepared students to be able to do the task. The average task score and average observation score were added to develop a final authentic pedagogy score (which could range from 7-30). Table 15 depicts the final authentic pedagogy (AP) scores along with demographic information associated with each teacher.

Table 15

*Teacher Profiles*

	Roy	Andy	Jason	Amy	Phillip	Lauren	Ryan	Lee
AP Score	9.6	10.9	11.6	12.9	13.3	18	20.9	21.2
Age	26-35	36-35	36-45	46-55	26-35	46-55	26-35	36-45
Ethnicity	White	White	White	White	White	White	White	White
Experience	4	11	14	15+	6	15+	11	12
Grade Taught	9	10	10	9	10	9	10	9

It should be emphasized that teachers in this study were not labeled as “authentic” or “traditional”. The authentic pedagogy scores represent a continuum. Teachers who scored at the high end of the continuum occasionally used strategies that would be considered more traditional (i.e. lecture; multiple-choice tests). However, the observation and interview data suggested that this type of instruction was not their dominant practice. These teachers seemed to have a fundamentally different conception of high level understanding than their peers. The next chapter will explore the extent to which this resulted in differences in student learning on the outcome measures.

When considering the first research question, I hypothesized that the mean score of the teacher sample would not reach the mean of the authentic pedagogy scale. I also believed, based on my purposeful selection of the research site to increase the likelihood of having some high scoring teachers, that enough variation would exist among the teachers in the sample to ascertain the impact of authentic pedagogy on student learning. My hypothesis was supported. The average authentic pedagogy score of 14.8 did not reach the midpoint of the authentic pedagogy scale which is 18.5. There was enough spread among the teachers to be able to address my other research questions. The final



authentic pedagogy scores were organized into four categories: minimal, limited, moderate, and substantial. The cut scores for these categories are listed in Table 16. The dividing points represent a breakdown of the totality of possible scores into approximate quartiles. They also correspond with those used by the broader Social Studies Inquiry Research Collaborative (SSIRC) study.

Table 16

*Cut Scores*

	Average Task	Average Instruction	Cut Offs
Minimal	3-3.99	4-8	7-11.99
Limited	4-5.99	8-12	12-17.99
Moderate	6-7.99	12-16	18-23.99
Substantial	8-10	16-20	Above 24

Some general statements and trends are evident based on an analysis of these data. First, when applying the cut scores to the authentic pedagogy scores in Table 15, it is evident that no teachers reached the highest “substantial” category of authentic pedagogy. This finding is not surprising given the difficulty associated with achieving the top levels of the rubrics. Three teachers, however, did score in the moderate range. These teachers were Lauren, Ryan, and Lee. They scored a good deal higher than the rest of the sample. Roy, Andy, and Jason were on the opposite end of the continuum in the minimal authentic pedagogy category range. The remaining teachers in this sample could best be characterized as using limited authentic pedagogy.

Lauren barely reached the moderate category with a score of 18. However, this score was likely influenced by circumstances surrounding the collection of her data.

Lauren retired during the study before I could observe the tasks she submitted. In order to document the degree of authentic pedagogy students experienced in her class, I rated her instruction based on two videotaped lessons that were a part of a previous lesson study project. During this project, Lauren created an inquiry-based lesson with the assistance of her 9<sup>th</sup> grade social studies colleagues as well as teacher educators and historians from Auburn University. As a result, the task scores associated with the videotaped lessons were very high. The three original tasks that Lauren submitted were not as authentic. It is likely that her final authentic pedagogy score was inflated. Nevertheless, Lauren represented a teacher who required students to complete some intellectually challenging tasks during the course of the semester.

The two highest scoring teachers in this study were formally trained in inquiry-based instructional practices. Lee had a master's degree in social studies education while Ryan earned his doctorate in the same field during the study. In addition to their formal training, Lee and Ryan also had extensive experience in applying this knowledge in the classroom. They were actively involved in inquiry-based professional development programs as participants, leaders, and mentor teachers. Their scores in this study suggest that a combination of graduate work and experience, filtered through a disposition that is amenable to the assumptions of inquiry based instruction, may contribute to higher levels of authentic pedagogy. These factors will be discussed further in the generalizations section of this chapter.

The teachers in the lowest category, by way of comparison, did not have the same extended experience with inquiry-based teaching. Roy and Andy held graduate degrees in fields other than social studies; administration and special education. Andy was a

veteran teacher who was actively transitioning into an administrative role during the study. He had attended at least one inquiry-based professional development workshop, but there was less evidence that he applied what he had learned in his teaching. Roy was the novice teacher in the study having taught social studies for only two years. Roy's understanding of inquiry-based teaching did not appear to be much different from some of the teachers in the limited category. However, his ability to implement this type of instruction was likely influenced by the growing pains associated with being a new teacher. Roy and Andy could best be described as teachers who predominantly used a traditional instructional approach. As Andy told me during one of our meetings, "I pretty much just lecture." Jason was also an experienced teacher who had over ten years in the classroom. However, most of this was in another state. In his two years at the study school, he had participated in at least one inquiry-based professional development workshop. Jason seemed comfortable using certain inquiry-oriented strategies, but still maintained a predominantly traditional instructional approach.

Amy and Phillip were in the limited authentic pedagogy category. Amy was a veteran teacher who had recently been involved in an intensive inquiry-based lesson study project. There was some evidence that she was becoming more comfortable and proficient in implementing problem-based lessons. Phillip was less experienced, but had more formal training in this type of instruction since he had graduated from an undergraduate social studies program that emphasized this approach.

The next section includes examples of the types of intellectual challenges students experienced at each of the levels of authentic pedagogy represented by this sample. The examples do not contain the level of detail associated with audio or videotaped

transcripts. However, they do help to explain the teacher's placement on the authentic pedagogy continuum. In each category (minimal, limited, moderate), at least one teacher's authentic pedagogy task scores are described to provide the reader with a general sense of the type of tasks that were submitted. This is followed by a more detailed explanation of a specific task/lesson combination which describes what students experienced and the scoring rationale.

### **Minimal Authentic Pedagogy**

Roy was the best example of a teacher who utilized minimal authentic pedagogy. Roy taught 9<sup>th</sup> grade World History to three blocks of students each day. The classes I observed typically averaged about 18 students. His students represented a broad range of ability levels to include some special education students.

Roy was a younger teacher (between 26 and 35), but not fresh out of college. His undergraduate teaching preparation was in social studies education. He later attained a master's degree in mild/moderate disabilities. Roy's first two years of teaching at the study school were in special education. He was completing his first year of social studies teaching when this research project was initiated. In addition to his teaching responsibilities, Roy served as a football and baseball coach.

In many ways Roy seemed to be still getting the feel for teaching. He was not as confident as his veteran colleagues in maintaining classroom discipline. His style tended to be inconsistent (of course this could have been due to my presence in the classroom). At times, he was overt and somewhat aggressive when addressing behavioral issues. In other situations, he was overly permissive with students. Routine administrative tasks and minor behavioral issues consumed a good deal of this teacher's time and energy.

This possibly played a role in his hesitancy to adopt a more student-centered classroom environment.

Table 17 depicts the tasks that Roy submitted and the final scores that were assigned using the task and instruction rubrics (Appendices C-F). Roy selected a political cartoon, an illustrated timeline, and an activity where students taught their classmates as his three tasks that most represented students thinking at a high level. The intent of the political cartoon activity, as explained in the interview, was for students to demonstrate a deeper understanding of some of the causes of World War I. Roy wanted students to show they understood complex terms like militarism and nationalism. The second task, the illustrated timeline, was part of a unit on the Industrial Revolution. In this lesson, students were asked to draw the larger significance of a list of events, inventions, people, or concepts associated with this time period. Finally, the “teach a lesson” task involved students teaching their classmates about events in a chapter on Nationalism. Roy believed the students would have to use higher order thinking to prepare an activity for the class to complete.

Table 17

*Overview of Roy’s Authentic Pedagogy scores*

Task Name	Task Score (3-10)	Instruction Score (4-20)	Final Authentic Pedagogy (7-30)
Political Cartoon	4	5	
Industrial Revolution Illustrated Timeline	6	6	
Teach a Lesson	4	4	
Average Task/Instruction Scores	4.6	5	9.6

Each of these tasks had the potential to be intellectually challenging. Political cartoons, for instance, can be used to convey complex messages and subtle nuances about a topic. With proper scaffolding, students can learn how to question cartoons like they would any historical artifact. A great deal of higher order thinking is often needed to uncover the potential bias of an artist or interpret the meaning of symbols. Teachers can lead students to question the artist's intent in including certain design features (i.e. color, symbols, etc.) or even have them create a cartoon expressing an opposing viewpoint. As part of an inquiry based activity, students might investigate the details surrounding an ill-structured problem or question. They might construct a cartoon that supports their view or the view of some historical group associated with the topic. In this case, the cartoon's message constitutes the basis of an argument – something the students can articulate as part of a class discussion or debriefing. Symbols and other features of cartoons are used to convey a message that has some real depth.

Roy's tasks did not live up to their description in the interview. Challenging tasks usually require detailed instructions and precise scaffolding to help students successfully think at a higher level. These features were noticeably absent in most of the assignments Roy provided to his students. In some instances, Roy seemed to confuse comprehension level tasks with higher order thinking. There was an overall disconnect between Roy's stated intentions and his ability to implement lessons which required the types of thinking envisioned by the authentic pedagogy model.

Roy's students had very little opportunity to complete assignments that required construction of knowledge or elaborated communication. The construction of knowledge scores for the three tasks (1,2,1 on a three point scale) suggest that Roy's dominant

expectation was for students to “reproduce information gained by reading, listening, or observing” (instruction rubric). This was especially true of the “teach a lesson” task that will be described later in this section.

Roy’s use of a political cartoon activity and illustrated timeline showed that he was willing to try alternative forms of assessment. Visual tasks such as these can require a great deal of elaborated communication on the part of students. However, Roy’s scores in this category were also fairly low (2,3,2 on a four point scale). His visual tasks were not as demanding as similar ones used by other teachers in this study. The political cartoon activity, for example, required students to draw a picture of one of the causes of World War I (a subject just covered in class). This doesn’t really constitute a political cartoon. Political cartoons usually involve attempts to persuade others or convey a point of view. The WWI drawings in the class that I observed were very simple. When students presented them, they usually could describe their idea in one or two sentences. Most of Roy’s tasks could best be described as the equivalent of a short answer exercise (level two on the task rubric).

Finally, all three of Roy’s tasks had virtually no connection to students’ lives (1,1,1 on a three point scale). The tasks did not require students to explore the modern relevance of historic events like World War I and the Industrial Revolution. They also didn’t provide students with much of an opportunity to study these topics in a way they would find personally meaningful.

The task that most exemplifies the type of instruction students received in this category was called “teach a lesson.” The scores for each standard of the task and

instruction rubric are provided in Table 18. Most of the standards received the lowest possible score.

Table 18

*Scores for “Teach a Lesson” Task*

<b>Task Scores</b>		<b>Instruction Scores</b>	
Construction of Knowledge	1	Higher Order Thinking	1
Elaborated Communication	2	Deep Knowledge	1
Connection to Students’ Lives	1	Substantive Conversation	1
		Connectedness to the Real World	1
Total Task:	4	Total Instruction:	4

Roy had used the “teach a lesson” activity at other times during the semester with different topics in the curriculum. The instructions he provided to students on this particular day were very brief (see Figure 3). The students worked on this task during the last day of a seven day unit. It was designed to prepare students for an upcoming test. Roy’s other classes did not do this assignment. He felt that it would work best with his first period; his strongest group of students.

**Teach a lesson**

You will be assigned a section from the Nationalism unit.

1. As a group you need to decide on 10 facts you and your classmates need to know. You need to write these 10 facts on your paper to present to the class.
2. Come up with an activity for the class to do. You must also create an example of your activity for the class to see.

*Figure 3.* The “Teach a Lesson” Task



When Roy implemented this task he organized students into five groups of three to four students each. The groups were assigned one of the sections from the textbook chapter they had been covering. Their task was to identify ten facts for their classmates to know. Once identified, the tasks were to be transferred to “cheerleader” paper for display during the presentation phase of class. In addition to identifying facts, each group was responsible for preparing an activity. The activity portion of the assignment could involve a demonstration of some sort by the group or a more interactive activity that involved the class.

Students worked on the activity for an hour. The groups divided up the task as might be expected. Those with better handwriting or artistic ability did the fact poster. The other group members thought of the activity or found the facts. In each group, it was usually one or two students that really looked through the materials (textbook, worksheets, notes) for the facts to incorporate into the presentation. This activity did not require students to justify why they chose certain facts as important. As I watched the lesson, the students mostly pulled sentences verbatim from their source. A girl seated near me opened up the textbook, identified the section, and looked at the sub-topics. She then said, “We can just do two from that, that, that, that, and that.” In this simple manner, the student selected two facts from each of the highlighted topics in the section. I asked a student from another group how she identified facts to include on the poster. She told me she was basically pulling the main points from the class notes. I viewed similar patterns from the other groups.

The talk within the groups was almost entirely procedural and often off topic. Each student had their own laptop computer. Several of the groups typed rough drafts

and then read the facts to the person writing on the butcher paper. This process was very inefficient when you consider that the class was equipped with an electronic whiteboard. Rather than use butcher paper, the files could have easily been pulled up on the whiteboard for everyone to see. Students seemed to sense the lack of urgency in the lesson and stretched out the task to take full advantage of the time provided by the teacher.

The presentations took up the final portion of class. Each group read their ten facts to the class. The first group played a game of “trashketball.” One student read multiple choice questions from a worksheet while the other two students in the group took turns answering. Approximately four questions were asked (related to the ten facts). A point was awarded for each correct answer. A correct answer also entitled the student to shoot a wadded up piece of paper into the trashcan. If the student made the basket, he/she got another point. The class looked on while the group played the game.

The second group also played a short game with multiple choice questions. This time, two questions were asked to the class. If a student got the question right, he/she got to “make a beat” (by banging on the desk). The third group had a word find that appeared to come from a textbook or perhaps online (not created by the students). The teacher showed it to the class. However, neither the class nor the group did anything with it. The fourth group located a map to support their facts. The teacher had one of the students explain the map. It showed the size and power of the Ottoman Empire. Finally, group five asked the class to write a half page diary entry on a serf. The class ended immediately after this presentation and there was no indication that the students were going to actually do this assignment.

The instruction rating for this task was fairly straightforward. It received a one on the higher order thinking scale because I did not notice any students engaged in higher order thinking during the lesson. The students simply took facts from their textbook, notes, or worksheets and transferred them to butcher paper. The activities created by the groups were often games that involved questions taken directly from worksheets.

Since the entire purpose of the activity was to help students memorize content for the test, the depth of knowledge standard also received a one. It clearly corresponded with the instruction rubric statement “students were involved in the coverage of simple information which they are to remember.” The students were not required to organize their facts into any sort of an argument to assist with learning.

The substantive conversation standard received a one because the conversation during class (that was on topic) was almost entirely procedural. I did not witness any instances of students grappling with the meaning of an idea or concept in the unit. They did not argue within their groups over which facts should be included in the presentation. The lesson also received the lowest score on the connectedness standard. This was a strictly “school” task. At no point did the teacher justify it beyond doing well in the course.

Other tasks in the minimal category provided a similar level of intellectual challenge to students. A comparison of Roy’s task with one submitted by Andy (see Figure 4) illustrates this point. Both teachers seemed to equate high level student understanding with activities that mainly required students to master large amounts of factual material. Roy’s “teach a lesson” activity had students pull ten facts from the book. Andy’s Reformer’s powerpoint asked students to describe specific information

related to reforms enacted in the 1800s. In each instance, the teacher used the tasks to help students learn nearly an entire chapter of information from the textbook.

Andy's task was a little more demanding than Roy's in the elaborated communication category simply because students had to include more information in their presentations. However, the quality of the information was essentially the same. The students were either summarizing information or listing facts from the textbook. They were not asked to form a generalization about the time period and back it up with supporting evidence. The task did not require analysis or persuasion and therefore fell short of the type of elaborated communication envisioned by the authentic pedagogy model. Finally, the task didn't score very well on the connection to students' lives standard since it was focused entirely on life in the 1800s. The final scores for this task were 1,3,1.

## Reformers of the 1800s

**Rationale:**

Students are to create a power point presentation that identifies 5 areas of the reform movement during the 1800s. This will give students an understanding of the social changes America experienced during the rapid growth of urbanization of the 1800s. Students will be able to better understand the concepts, developments, and consequences of industrialization and urbanization.

**Procedure:**

You are to identify 5 areas in which significant reform occurred during the 1800s and create a 20 slide power point presentation. In each area identify the most significant leaders, 3 supporting facts or reasons for the reform, and the legislation that was enacted because of the reform.

**Set Up:**

Slide One-Name of Reform

Slide Two-Leaders of the Movement

Slide Three-Supporting facts/reason for the reform

Slide Four-Legislation

**Due Date:**

TBA. This should be delivered to my R drive and also be located on the student's P drive.

---

**Rubric for Assignment**

Name \_\_\_\_\_

This assignment is worth 100 points as a major test grade

<b>CATEGORY</b>	<b>POSSIBLE PTS.</b>	<b>POINTS AWARDED</b>
Set up Criteria	40	_____
Accuracy	40	_____
Timely Delivery	10	_____
Creativity	10	_____
<b>TOTAL</b>	<b>100</b>	_____

*Figure 4.* Reformers of the 1800s Task

## **Limited Authentic Pedagogy**

The teachers in the limited authentic pedagogy category generally had higher task scores than the teachers in the previous category. However, their instruction scores remained relatively low. They struggled to provide the support needed for students to accomplish their higher order thinking goals. This section uses Amy's tasks as the basis for understanding the types of intellectual challenges students experienced in classrooms featuring limited authentic pedagogy.

Amy was a veteran teacher with over fifteen years of teaching experience. She had taught World History at the study school for five years. Her professional training was in elementary education. The undergraduate and graduate programs she completed provided enough social studies credits for her to be certified to teach in this field.

As might be expected, Amy's classroom had a different feel from Roy's in terms of basic management. Amy was an excellent classroom manager who exhibited a no-nonsense approach to instruction. All three of her tasks involved substantial groupwork and movement of students within the class. She was able to seamlessly transition through the different stages of these lessons with little difficulty. The classroom atmosphere was relaxed, yet focused. Her students seemed to genuinely enjoy coming to class.

Some veteran teachers settle into a teaching routine and become resistant to ideas that challenge their status-quo. Amy was not this type of teacher. She sought out opportunities for professional development and growth. Amy was involved in the same inquiry-based lesson study project as Lauren and some of the other teachers in this study. The teachers in this project worked closely with a professional historian to develop in-

depth content knowledge of several major topics from the World History curriculum. They then used this knowledge to prepare inquiry-based lessons.

In observing Amy’s instruction, it was evident that she knew her subject very well. It was also obvious that she had incorporated specific strategies from the lesson study project into her instruction. However, the tasks she submitted for this project suggested that her adoption of an inquiry-based approach was still mostly limited to the lessons she had developed with her peers. Table 19 provides an overview of the three tasks she submitted and how they scored on the task and instruction rubrics.

Table 19

*Overview of Amy’s Authentic Pedagogy scores*

Task Name	Task Score (3-10)	Instruction Score (4-20)	Final Authentic Pedagogy (7-30)
Absolute Monarchy of your Own	4	8	
Ideal Form of Government Debate	9	10	
Renaissance Ball	4	4	
Average Task/Instruction Scores	5.6	7.3	12.9

Amy had two lower scoring tasks and one that was significantly higher which helped her overall authentic pedagogy score to extend into the limited range. The first task she submitted was called an “Absolute Monarchy of your Own”. The purpose of this task was to reinforce students’ understanding of the term “absolute monarchy” and how one would likely function. Amy’s intent was to have students synthesize what they had learned about various absolute monarchies in history into a fictitious example that they could relate to personally. She also wanted students to evaluate absolute monarchies and the effect they could personally have on others. Students worked in groups to create a

fictitious kingdom where they were the absolute ruler. The handout for this activity required students to explain how their kingdom would function (i.e. how will you direct your subjects to worship?). This allowed them to see the type of power a king would actually have under this system.

Amy's second task was a debate where students represented the views of different philosophers (Locke, Plato, etc.) and discussed the ideal form of government. At the end of the debate, they had to step out of their assigned roles to argue for the form of government they considered the best. The final task was another perspective taking exercise where students assumed the role of a Renaissance figure to participate in a Renaissance Ball. Amy wanted students to empathize with the historical figures and what it was like to live during this time period. The activity included an initial "meet and greet" session where students got to know the cast of characters at the Ball. Then, students settled into their seats and were called on by the teacher to discuss their greatest accomplishments using various props to support their presentations.

At first glance, the tasks submitted by Amy appeared to differ significantly from those of Roy or Andy. They certainly required a good deal of active participation and engagement on the part of the students. However, two of the tasks were essentially creative alternatives to lecture which required little construction of knowledge (1,3,1). The rubric for the absolute monarchy assignment provided the greatest insight into the teacher's expectations for this task. In order to get full credit, students simply had to follow directions and include each of the required elements listed on the assignment handout. The task itself was not very intellectually challenging because students could develop their monarchy in any way they deemed appropriate (which admittedly was the



point). It was perhaps challenging from a creativity standpoint, but it did not require the disciplined use of higher order thinking processes to solve a problem. Amy attempted to address the synthesis and evaluation goals for this assignment during the lesson and this is reflected in her instruction score.

The Renaissance Ball task mainly involved students reporting factual information about their character. The students were not trying to master the information in order to formulate an argument related to a problem or central question. The manner in which this task was implemented suggested that its dominant purpose was to help students remember the main achievements of each Renaissance figure in order to perform well on the upcoming unit test. The final task, the Ideal Form of Government, did involve substantial construction of knowledge and will be discussed later in this section.

Amy's elaborated communication scores were also fairly low (2,4,2). The Absolute Monarchy and Renaissance Ball tasks required students to do a great deal. For instance, as part of the Renaissance Ball, students prepared a short poem, a mask of their Renaissance figure, and a bust of their figure with accomplishments listed on it. They also designed props for their presentations and in some case wore costumes. However, the elaborated communication standard isn't as concerned with how much students do, but rather the extent to which they are required to explain and defend their understanding of historical concepts. Both of these tasks elicited very brief responses from the students in Amy's class. In most cases, the students answered the questions in a couple of sentences.

Finally, most of the tasks did not require students to connect what they had learned to something significant in their lives (1,2,1). Why might it be important to

understand absolute monarchies or the lives of Renaissance figures? It was likely that some students recognized the contemporary relevance of these topics, but the tasks themselves did not press them to investigate these connections in any detail.

Amy’s highest scoring task featured a debate on the ideal form of government. The class I observed was fairly evenly divided among boys and girls and included a total of sixteen ninth grade students. The students were mostly white (11/16) and considered regular education in terms of their ability. The task was introduced near the beginning of a unit on the Enlightenment. The students prepared for the debate during the course of several class periods. I observed the debate itself on the last day of the unit. Table 20 provides a breakdown of the authentic pedagogy scores associated with this task.

Table 20

*Scores for “Ideal Form of Government” Task*

<b>Task Scores</b>		<b>Instruction Scores</b>	
Construction of Knowledge	3	Higher Order Thinking	3
Elaborated Communication	4	Deep Knowledge	3
Connection to Students’ Lives	2	Substantive Conversation	3
		Connectedness to the Real World	1
<b>Total Task:</b>	<b>9</b>	<b>Total Instruction:</b>	<b>10</b>

The teacher’s intent was for students to learn the views of nine historic thinkers on the ideal form of government and their beliefs regarding the role people should play in governing. After considering the various perspectives, students were to evaluate the different forms of government in order to decide which one they considered to be the best. Amy had the students defend their choice through an editorial assignment. The

ideal form of government task is based on a *History Alive* activity. Amy's editorial was added in place of the debriefing.

Amy set up the debate so that students were arranged in a two-deep semi-circle with "actors" seated in front of their "press agents". The actors wore paper masks resembling the historic philosopher they were attempting to portray. Some also wore togas. They had nameplates on their desk for easy identification. Due to absences, Amy allowed some of the more academically able students to not have a press agent. At the beginning of class, the teacher passed out a data retrieval chart for students to complete during the debate. The press agent's responsibility was to introduce their assigned thinker. Each agent delivered a prepared statement with pertinent background information for their actor. During the debate, the press agents really did not participate other than taking notes.

Amy served as the moderator of the debate. She called on one of the actors playing an historic figure, their press agent would do the introduction, and then the actor would explain the symbol on his/her nameplate. The symbol had to represent their views on the ideal form of government. Amy would then ask for questions. After a few questions and some discussion, she would move to the next historical figure and repeat the process. The discussion surrounding each philosopher was usually 8-10 minutes in duration. The following dialogue provides a general sense of how this lesson was implemented and the type of discussion that took place. It is not a verbatim transcript, but it does capture the essence of what was said during the discussion.

Teacher: Any more questions? Alright Hobbes. Your press agent isn't here today.

Can you tell us a little bit about yourself?

Student playing Hobbes reads his biography from a card.

Teacher: What is your symbol?

Hobbes: People with crowns on their heads – governing themselves. I resent that.

Teacher: What is the ideal form of government? Why do you consider it ideal and can we trust people to govern themselves?

Hobbes: The ideal is an absolute monarchy. People can't govern themselves.

Teacher: Why?

Hobbes: People are selfish. They are out for their own selfish interest.

Plato: Doesn't that contradict your previous statement?

Teacher: What do you mean?

Plato: The absolute monarch is a person. Could he be corrupt himself?

Hobbes: One person is no big deal. If everyone is corrupt and in charge, things are worse.

John Locke: But wouldn't he act selfishly?

Hobbes: (hard to understand his response – seems to be trying to understand the question)

Locke: If the absolute monarch is selfish and corrupt – how would that work out?

Hobbes: If one person is in charge, even if he is corrupt – it is still better than people governing themselves.

Plato: Why do you think people are so corrupt?

Hobbes: People have to carry guns and lock their doors.

Plato: I don't carry a gun or lock my door.

Hobbes: the high rate of crime.

Wolstonecraft: The absolute monarch is better than the people so he should rule – is that what you are saying?

Hobbes: (mumbled response – teacher had to ask the student to speak up)

Rousseau: How do you believe in passing on a monarchy? Would that be fair?

Hobbes – (response isn't clear)

John Locke: Everyone is corrupt. That is correct?

Hobbes: Yes

Locke: Does that include you?

Plato: You seem to put a lot of trust into one person.

Hobbes- keeps saying the same thing – one ruler is better than everyone governing themselves.

Locke: Uses a tug of war analogy. He suggests that people pulling on both ends would achieve a position in the middle (negating some of the corruptness) whereas if there is only an absolute monarch on the line there would be nothing to counter his selfishness.

Teacher: Hobbes is staying true to his beliefs despite much controversy. Any questions? (a few more comments that are similar to the previous ones)

Teacher: O.K. –whether we agree or not, I think we understand Hobbes' position.

The full segment lasted for approximately eight minutes. Much of this time was spent with students trying to get Hobbes to justify why an absolute ruler could be trusted to lead. This excerpt illustrates the redundant nature of some of the dialogue. It also shows how some students struggled to fully grasp their character's perspective. In this case, the student playing Hobbes had a basic understanding of his beliefs, but when

pressed to defend his position he was unable to discuss factors that made his rule legitimate (i.e. divine right). He also had a difficult time supporting his statements with historical evidence. Other students were more proficient. The teacher allowed students to freely debate with each other. This made the debate feel more authentic.

Most actors accurately portrayed the point of view of their historic figures. The students were respectful towards each other and some genuinely seemed interested in trying to understand the perspective of the other thinkers. Certain students dominated the question/answer period and were obviously more knowledgeable about the subject. In particular, the student playing John Locke succeeded in raising some significant questions that led to some higher order conversation. A debriefing which allowed students to step out of character would have been helpful in enabling the class to come to some shared understandings of what these different philosophers believed. This would also allow the teacher to discuss some of the interesting comments made during the debate and address any misconceptions prior to the editorial assignment. Amy seemed to expect students to be able to synthesize and form important connections from this lesson on their own. She may have underestimated the complexity of the activity and the level of scaffolding needed to help students accomplish its higher order goals. Another possibility is that, having worked with these students for a while, she may have formed preconceptions regarding their abilities. Amy might not have realized how powerful a good debriefing could be in helping students, from a range of abilities, to achieve at a higher level. Overall, the class seemed to really enjoy the assignment and most students were paying attention even if they didn't contribute to the conversation.

When scoring the lesson, I worked from the bottom level of the instruction rubric to the top. The difference between levels, particularly for a three or higher, often comes down to a numbers game. In order to assign a “five” to a standard, I had to observe “almost all” of the students engaged in the desired behavior (i.e. higher order thinking, substantive conversation). Some categories, such as substantive conversation, are easier to score because they are overt behaviors. The higher order thinking standard is probably the most difficult. During the ideal form of government debate, I had to listen closely to student comments for indicators that higher order thinking was taking place (i.e. synthesis, evaluation, analysis). It was likely that comments made during the debate provoked higher order thinking in members of the audience (i.e. the press agents), but they didn’t have the opportunity to make any comments during the lesson. My scoring on this standard and the others was limited to what I could physically observe.

The Ideal Form of Government lesson received a three on the higher order thinking standard because I did not observe “many” students engaged in higher order thinking; a requirement for the next level. The rubric defines “many” as at least 1/3 of the class. In this case, the 1/3 standard required roughly five of the sixteen students to demonstrate higher order thinking. This standard was difficult to accomplish because the press agents didn’t participate in the debate. The omission of the press agents left nine remaining students on the panel of historic characters. Three of these students provided little input during the debate beyond their own presentations. The six panelists who were active throughout the debate demonstrated higher order thinking sporadically at best. Most of the opening statements were scripted. The higher order thinking mainly occurred when students asked original probing questions to uncover weaknesses in another

character's argument or when they justified their own position on the ideal form of government. The criteria for a three, "some students perform some HOT operations", best describes what I observed in this lesson.

The depth of knowledge standard also received a three. This standard measures the extent to which students achieve a nuanced understanding of the lesson content. A lesson can also score relatively well if the teacher demonstrates deep knowledge. A level three score indicates that knowledge was treated "unevenly" during the lesson. During the ideal form of government debate students demonstrated deep knowledge in some areas, but only superficial understanding in others. Most of the students seemed to be able to describe the key differences between the various forms of government featured in the debate. They also seemed to understand the idea of divine right to rule. However, it was evident that the students playing Montesquieu, Hobbes, Plato, and perhaps others did not have a deep understanding of their character. They could not address questions that required them to go off script and make generalizations based on information from their research.

High scoring lessons in the depth of knowledge category must also maintain a sustained focus on a significant topic. This lesson initially seemed to meet this criterion since it was oriented around an important central question. However, the bulk of the class period was spent exploring the views of the different philosophers. The debate was segmented to ensure enough time was allocated to hear from all of the actors. As a result, the students had a limited opportunity during class to synthesize and apply their knowledge to the central question. The lesson could have reached a four if the teacher



took a more active role in asking probing questions and challenging students' to defend their point of view at some point during the lesson.

I also assigned this lesson a three for substantive conversation. A level four requires all elements of substantive conversation to be present (sharing, coherent promotion of collective understanding, higher order thinking). The coherent promotion of collective understanding aspect of the standard was missing in this lesson due to the fragmented nature of the debate (moving from philosopher to philosopher with no free debate out of character). The students didn't really form any conclusions on the central question during the class. The lesson definitely featured sharing of ideas between students and at least one example of sustained conversation which is defined as at least three consecutive interchanges (a statement by one person and a response by another). As a result, it best fit the requirements for a level three score.

The final scoring category is value beyond school. Amy never provided any justification for studying the philosophers' views on government. According to the instruction rubric, in a class with little value beyond school, "activities are deemed important for success only in school (now or later), but for no other aspects of life. Student work has no impact on others and serves only to certify their level of competence or compliance with the norms and routines of formal schooling." The editorial assignment seemed to fit this description; certifying competence according to the norms of formal schooling. Students may realize this lesson has value outside of school, but they didn't verbalize this understanding during class and as a result I assigned this category a one.

The big difference between the teachers in the limited category and the moderate category was in the instruction students experienced. Teachers in the limited category developed some challenging tasks, but they were not as successful in helping students accomplish their objectives. Phillip's tasks provide another example of this overall theme. His authentic pedagogy scores are summarized in Table 21. Phillip submitted a document analysis task, an 18<sup>th</sup> century reformer's task similar to Andy's, and a painting analysis. The first task was an analysis of George Washington's Farewell Address. Students read his address and answered comprehension questions. The class then discussed Washington's views regarding foreign alliances and political factions. After the discussion, the students read an article about an Iraq war appropriations bill being considered by Congress. The concluding discussion focused on whether the United States was doing a good job of heeding Washington's advice today.

Phillip's second task had students research a portion of the textbook chapter dealing with reformers from the 1800s. Students had to develop a powerpoint presentation describing the three most significant ways an individual or event contributed to reforming America. The central question associated with this task asked: Did the era of reform serve to better America in ways that are still represented in today's United States?

The final task was an analysis of the painting "American Progress" by John Gast (see Appendix P). Phillip used this painting to help students understand the concept of Manifest Destiny. The class examined positive and negative consequences associated with America's expansion during this time period.

Table 21

*Phillip’s Authentic Pedagogy Scores*

Task Name	Task Score (3-10)	Instruction Score (4-20)	Final Authentic Pedagogy (7-30)
Washington’s Farewell Address	8	8	
Reformers Lesson	7	4	
Manifest Destiny Painting Analysis	6	7	
Average Task/Instruction Scores	7	6.3	13.3

Phillip’s tasks scored at the mid-point of the construction of knowledge standard scale (2,2,2 out of 3). Each task included at least “some expectation for students to interpret, analyze, synthesize, or evaluate information, rather than merely to reproduce information.” The Manifest Destiny Painting Analysis task included a series of questions that students were to answer in order to better understand the artist’s perspective on the time period. One of the more challenging questions called for students to compare Gant’s painting to Emmanuel Leutze’s “Washington Crossing the Delaware”, a painting they had studied earlier in the semester. The task probably would have reached a level three if the questions more consistently called for students to defend their responses. The document analysis task also involved at least some construction of knowledge. Students had to apply their knowledge of Washington’s address to a contemporary foreign policy issue. The final task based on the 19<sup>th</sup> century reformers barely met the standard for a two in this category. Phillip’s intent was for students to develop an argument justifying the significance of their reformer.

Each of Phillip’s tasks scored near the top of the scale in the elaborated communication category (3,3,3). His tasks required students to explain their

understanding of historical concepts in ways that exceeded short answers or one word responses. In most cases, the students had to provide a short summary of their conclusions. The limited space for student responses on the questioning scaffolds for the document analysis and painting analysis provided the best indication of the teacher's expectations. The tasks did not require extended responses where students had to make generalizations and support them with evidence.

The final category was connection to students' lives. Phillip's scores spanned the entire range of the scale (3,2,1). His highest scoring task was the analysis of Washington's Farewell Address. This task had students consider issues of contemporary relevance; the influence of political factions and U.S. involvement in foreign alliances. Students expressed their own views on these topics while considering whether the U.S. was doing a good job of following Washington's warnings. The Reformer's lesson also provided students with at least some opportunity to consider the modern significance of reforms enacted in the 1800s. The lowest scoring task in this category was the painting analysis, which was situated entirely in the past.

In summary, the tasks associated with the limited authentic pedagogy category were typically more ambitious than those from the minimal category. They were more likely to elicit at least some higher order thinking from the students. In implementing the tasks, the teachers sometimes struggled to maximize their instructional value. Important learning opportunities were missed for a variety of reasons (i.e. inadequate scaffolding, the absence of a debriefing). The teachers seemed open to engaging students in inquiry, but had not reached the level of expertise of their peers at the moderate level.

## **Moderate Authentic Pedagogy**

The final authentic pedagogy category featured three teachers from the sample. The tasks submitted by these teachers often achieved the highest levels of the construction of knowledge and elaborated communication standards. Students were engaged in activities that required higher order thinking and significant writing designed to argue, convince, or persuade rather than just summarize or report information. Lee and Ryan, in particular, seemed to effectively couple these tasks with rigorous instruction. It takes a great deal of skill to exceed a level three score for any category on the AIW instruction rubric. These teachers routinely received four's and Ryan was the only teacher to achieve the maximum score for any of the authentic pedagogy standards.

Ryan's authentic pedagogy scores were fairly representative of this category. As a result, I will focus on describing what students experienced in his class. Ryan was a white, male teacher in his mid thirties. His professional degrees were in general social science education. During the study, he achieved his doctorate. Ryan had over eleven years of experience and was responsible for teaching both regular U.S. History and Advanced Placement European History classes. He also had experience teaching undergraduate classroom management and social studies methods courses.

Like Amy, Ryan was very effective in managing the learning environment. This went well beyond avoiding disruptions and ensuring students were on task. Ryan's classroom conveyed his passion for learning and discovery. It was essentially a miniature library. Ryan's desk was surrounded by stacks of books covering a wide range of social studies topics. Another set of books spanned the row of desks situated along the entire back wall of the classroom. The overall classroom environment suggested that the

teacher was probably well read (most of the books were his) and the students used more than just the textbook to understand the past. This was confirmed during one observation when I watched students use these books to look up information. Ryan's classroom environment (in and of itself) likely triggered the intellectual curiosity and interest of at least some students.

In general, Ryan was very good at motivating students to think. His students appeared to love his lectures because they included a good mixture of humor and sarcasm. However, they also required significant student involvement and discussion. Ryan pushed his students to consider more than just the facts. Ryan asked the hard questions and required his students to clearly explain their thinking. His discussions were also demanding. He acted as a true facilitator, allowing students to do most of the work while occasionally intervening to ask probing questions or shift the conversation in a new direction.

Ryan submitted two tasks for this study which were used primarily with advanced placement students. The first task was a think aloud activity where students assumed the role of Czar Nicholas during World War I. The students read a document (the think aloud) which provided a monologue of Czar Nicholas considering Russia's problems and various options at his disposal. The students used this document to formulate a realistic decision for improving the situation facing Russia in 1916. The second task was a World War II political cartoon analysis. The cartoon featured the major leaders of WWII seated around a dominoes table (Roosevelt, Churchill, Stalin, Hitler, Mussolini, Tojo). The position of the leaders and the dominoes on the table suggested that the Allies were winning the war and the Axis powers were nervous (See

Appendix Q). The class analyzed the cartoon and then the students had to write an original dialogue featuring all of the leaders. Each student was assigned a writing prompt from either an Allied or Axis perspective. The students had to comply with a number of requirements in writing the dialogue.

The third task was used in both AP and regular classes. Students completed the “Me Card” task during the first few days of school. The task required students to design a three by five card that answered the question: How should the world see me? The card could include virtually anything (i.e. collage clippings, origami, pictures, quotes, etc.). The students also had to answer eight sets of follow-up questions that covered a wide range of topics (i.e. Is the card a primary or secondary account?; Would the collection of these cards be an accurate depiction of this class?). The students presented the cards in class and then Ryan led a debriefing using the follow-up questions as a guide. The discussion introduced students to some of the challenges associated with interpreting the trustworthiness of historical artifacts. The task exposed students to the epistemological foundations of the discipline.

The authentic pedagogy scores for these tasks are listed in Table 22. The three tasks scored very well on the construction of knowledge scale (3,3,2 on a three point scale). Tasks that predominately engaged students in higher order processes such as interpretation and evaluation of information received the highest score on this standard. The Czar Nicholas think aloud was a three because students had to evaluate a situation and determine a solution to a historic problem. The political cartoon analysis also placed significant higher order thinking demands on the students. Students had to assume a particular perspective (Axis or Allied), synthesize relevant factual information, and

generate a plausible dialogue addressing a significant issue from WWII. The students had to understand and accurately represent competing views in order to include all of the leaders in the dialogue. The only task that didn't achieve the maximum score for the construction of knowledge category was the Me Card. This assignment featured some interpretation and synthesis of information. However, in order for a task to reach a three, it must force students to consider the nuances of a topic beyond surface level exposure or familiarity. This task was part of an introductory lesson for the semester. It was designed to simply introduce students to some significant historical thinking concepts. Ryan intended to build on this knowledge as the semester progressed.

Table 22

*Ryan's Authentic Pedagogy Scores*

Task Name	Task Score (3-10)	Instruction Score (4-20)	Final Authentic Pedagogy (7-30)
Czar Nicholas Think Aloud	8	14	
Political Cartoon Analysis	8	12	
Me Card	7	14	
Average Task/Instruction Scores	7.6	13.3	20.9

The scores for the elaborated communication category were similar (4,4,2 on a four point scale). The first two tasks required significant writing. In order to achieve a four, a task must call for generalization and support. In the first task, students were presenting an argument regarding what the Czar should do to improve Russia's situation using historical evidence as support. The WWII dialogue also went beyond just reporting or summarizing information. The students had to ground their interpretation of the focus



question/prompt in factual information and details from the time period. The “Me Card” assignment could best be described as a short answer exercise which fits the criteria for level two.

Ryan’s authentic pedagogy scores were lower in the connection to students’ lives category (1,1,3 on a three point scale). The problems or questions associated with the first two tasks were not the type that students are likely to encounter in their own lives. Students won’t have to figure out a way to save Russia in 1916. These tasks fit the criteria for a level one score because they “offer very minimal or no opportunity for students to connect the topic to experiences, observations, feelings, or situations significant in their lives.” In the cartoon task, the students were writing entirely from the perspective of the WWII leaders. The same was true for the Czar Nicholas task. Students might find a way to personally relate to the Czar’s circumstances, but the task did not require it.

The Me Card task was designed to promote a personal connection with the student and to help students understand how historical thinking skills might apply today. Significant elements of the students’ lives were used as the basis for this activity. One question in particular had students evaluate the trustworthiness of information on the cards. This was something that students will have to do in their daily lives.

I’ve highlighted Ryan’s WWII political cartoon analysis task as an example of a lesson in the moderate authentic pedagogy category. Scores on this task are provided in Table 23. This task was implemented with an Advanced Placement European History class. The class was relatively small. Sixteen students were present on the day I

observed. The students were predominately white and female. However, some Asian and African American students were also in the class.

Table 23

*Scores for “WWII Political Cartoon Analysis” Task*

Task Scores		Instruction Scores	
Construction of Knowledge	3	Higher Order Thinking	3
Elaborated Communication	4	Deep Knowledge	4
Connection to Students’ Lives	1	Substantive Conversation	4
		Connectedness to the Real World	1
Total Task:	8	Total Instruction:	12

The political cartoon task was implemented after students had already received a good deal of instruction on World War II. They had covered America’s initial entry into the war and the major events through the victory in Europe. Ryan intended to discuss the war in the Pacific after the cartoon activity. The class began with students taking a practice Advanced Placement test. The test lasted for approximately thirty minutes. When all the students were finished, Ryan went over the answers and provided the students with some test taking tips and strategies. The next portion of class focused on the political cartoon analysis activity. The cartoon analysis lasted for nearly half an hour. At the beginning of the analysis, the class attempted to determine the cartoon’s source, its context (time frame), and its bias (Axis or Allied?). Ryan listened to initial ideas and then the class examined the details of the cartoon more closely. The students identified all of the World War II leaders seated at the table in the cartoon. They argued about specific elements of the drawing (i.e. is that a bead of sweat on Mussolini’s forehead or a strand of hair?). Once students identified what was being portrayed in the cartoon, Ryan

pushed students to analyze the details in greater depth. For instance, he asked why the artist decided to use dominoes instead of cards. He also had students consider the positions of the various leaders in the picture (why is Stalin standing behind Churchill?). Each of these questions elicited significant discussion. Ryan emphasized that nothing was an accident in a created piece of artwork.

The use of a heuristic (Source, Analyze, Contextualize, Corroborate, Think Deeply) helped guide the discussion and prevent it from proceeding in a random, linear fashion. Instead it was more recursive. The students took an initial guess at whether the cartoon was public or private, the time period, and what the artist was attempting to convey. Then, they examined elements of the cartoon more closely. As they understood more of the symbolism, they returned to aspects of the heuristic and re-examined earlier comments. The class understanding of the cartoon built throughout the activity. Ryan guided the discussion by asking questions and modeling the analytic process. When student comments were brief, he asked follow-up questions to force them to elaborate and support their opinion. The following brief dialogue regarding whether the cartoon was public or private illustrates this point:

Student: Public.

Teacher: convince me.

Student: The drawing isn't very exceptional. It doesn't look good enough for someone to have commissioned it.

Teacher: O.K. - the old copy of the drawing hurts. It does look grainy.

Student2: Maybe it was published in a newspaper. That might explain the poor quality of the image.

Teacher: Does this look like something you might expect to find in a newspaper?

Several students agree and provide reasons to support their opinion.

Teacher: We touched on the point of view or possible bias in this picture earlier (how Hitler was being depicted). Public seems to be a good guess.

Student comments during the cartoon analysis were often more than just a couple of words. The analysis felt more like a true discussion. While Ryan was clearly in charge, the student comments reflected sensitivity to the ideas of others. For instance, one student began a statement about why she thought an element of the cartoon represented Churchill's last gamble by saying "I'd like to add to Maggie's comment about the U.S. holding all the dominoes." At another point in the analysis, a student admitted she was confused about part of the cartoon and several students explained it to her further. This type of student to student interaction was more commonly observed in Ryan's classes than the other classes I observed from this sample of teachers.

Once most of the major ideas in the cartoon had been teased out, the teacher introduced the writing assignment. The students were to complete a dialogue based on the scene in the political cartoon. Ryan described it as a movie short (imagine a camera that zooms from face to face). The students were assigned a particular perspective (Axis or Allied) and a central question to address. Students worked on this assignment for the remainder of the class period.

This political cartoon task was similar to the Manifest Destiny painting analysis task (from the limited category) taught by Phillip. Both activities involved the analysis of visual media. Ryan perhaps had some advantages when he implemented the cartoon analysis lesson. His students seemed motivated to practice a skill they would use on the

AP exam. They also had more background knowledge at their disposal since they experienced the activity later in a unit. Taking this into account, Ryan's lesson still seemed more effective. This was mainly due to the way he used scaffolding to support student inquiry.

Ryan's method of using open-ended discussion, guided by a heuristic, seemed to engage the students more than the painting analysis scaffold used in Phillip's class (see Appendix P). Phillip's scaffold was used more as a worksheet to record answers. Phillip's class never really analyzed the context of the *American Progress* painting. Phillip told the students it was painted in 1870 by John Gast. The students never tried to figure out who John Gast was or his motives for painting the picture (was it commissioned?, etc.). They didn't discuss the time period of the painting and how it might have influenced Gast's worldview. Ryan's heuristic seemed to help students develop a deeper interpretation of the cartoon. In Ryan's class, sourcing and contextualization was a central feature of the discussion. In Phillip's class, they were largely omitted.

The cartoon analysis was a fairly strong example of authentic intellectual work. The task received eight out of ten possible points. The instruction score was also above average (12 out of 20 points). It probably would have scored higher if the first part of class was not dedicated to the practice exam. A review of the instruction scores provides a better sense of what made the lesson score so well. The first standard is higher order thinking. I assigned this lesson three out of five possible points. A level three score indicates that the majority of the class was spent with students engaged in lower order thinking, but one significant question caused some students to engage in higher order

thinking. The cartoon analysis fits the description for this level since higher order thinking was more than just a minor diversion in the lesson. The class worked together for an extended period of time to apply the heuristic and determine the meaning of the cartoon. They applied the knowledge they had learning previously in the unit to a new task. In doing so they utilized a range of higher order operations. The students were able to deduce the time range of the cartoon based on nuances in the picture (i.e. Mussolini is still there, so it can't be later than 1943...). The students also engaged in higher order thinking when they tried to determine the artist's motives (i.e. why did the artist use dominoes instead of cards?) and perspective (was this published in Britain or the United States? Why?).

In order to assign a level four score, many students must be engaged in higher order thinking for a substantial period of time (at least 1/3 of the lesson). The cartoon analysis by itself lasted for 26-27 minutes. The AP exam and writing activity made it impossible for this lesson to meet the substantial portion threshold. It is possible that these activities evoked at least some HOT, but this could not be observed.

The lesson received a four on the depth of knowledge standard. It clearly met the criterion for this standard. The students analyzed a complex political cartoon which required a nuanced understanding of events from WWII and the ability to recognize the perspectives of the major leaders. The analysis was sustained for a significant period of time. Many students were actively engaged in this activity (at least 6 of the 16). Most importantly, they were doing most of the work. Ryan resisted the temptation to give students his own interpretations. The students formed reasoned, supported conclusions about the meaning of the cartoon with limited guidance and support. The follow-on

movie short activity was a perspective-taking exercise which also required a great deal of background knowledge and the ability to empathize with the views of historical figures. This standard did not reach a five because "almost all" of the students had to demonstrate depth of knowledge. This is very difficult to achieve. As in most classes, certain students dominated the discussion. These students seemed to have a strong understanding of the content, but I was less certain about the others. I could not state with confidence that 14 out of the 16 students' reasoning during the main activity reflected "fullness and complexity of understanding."

As previously stated, Ryan was particularly good at leading productive discussions. This was reflected by the substantive conversation score (4 out of 5) for this lesson. Many students actively participated in the cartoon analysis and the dialogue writing activity. I witnessed the sharing of ideas, students making distinctions and building off the comments of their peers, and higher order thinking. Like the depth of knowledge standard, this lesson did not reach a five due to the "almost all" students requirement.

The main weakness in this lesson was the connectedness to the real world standard (1 out of 4). This means the lesson didn't have a clear connection to anything beyond school. The most obvious connection for these students was the activity's relation to the AP exam. It was evident that the activity was good practice for analyzing visual media. The score on this standard could have been improved if Ryan connected the ability to interpret WWII era cartoons to effective citizenship perhaps by providing examples of how visual media is currently used to shape public opinion. The themes

embedded in the task also represent some persistent problems that could have been mentioned.

In summary, Ryan's cartoon analysis task resulted in students creating an original dialogue among the WWII leaders. Students successfully analyzed and interpreted the meaning of the cartoon and responded to some of the major themes associated with WWII (i.e. why didn't the U.S. enter the war sooner?). The lesson required elaborated communication both orally and in writing. The task didn't afford students much of an opportunity to form a personal connection to the content and its relevance to life beyond the classroom was not explicitly established. Nevertheless, it was still quite challenging and Ryan was able to effectively structure the lesson so students were able to meet his expectations.

Lee's "Truman Think Aloud" provides another strong example of the type of intellectual challenges students experienced in classes with moderate authentic pedagogy. This task was a perspective taking exercise where students attempted to get into the mind of President Truman during a pivotal event of the Cold War (See Appendix R). It involved an in-depth analysis of the Berlin Crisis. Students analyzed four historically authentic courses of action for dealing with this event and ultimately had to assume the role of President Truman to decide the best way to resolve the crisis.

The think aloud reached the highest level of the construction of knowledge standard (3) on the task rubric. The dominant expectation was for students to analyze an historic problem and evaluate the various options available to President Truman. The students ultimately had to present an argument representing the best possible approach



for resolving the Berlin crisis. Lee provided students with ample materials and resources to be able to formulate a deep, nuanced position.

The think aloud also scored well on the elaborated communication standard (4 on a scale of 4). Students had to write a speech explaining their solution to the Berlin crisis and in doing so they were making claims and supporting them with evidence from the various advisors. The task clearly exceeded a fill in the blank or short answer activity. It was also more than a report or summary. Students were engaged in writing meant to “convince or persuade” others.

The only area where the task didn’t score well was in the connection to students’ lives standard. The think aloud did not explicitly require students to discuss the modern significance of historical events. The scenario was entirely situated in the Cold War time period. Table 24 depicts the scores Lee received on both the task and instruction rubrics.

Table 24

*Scores for “Truman Think Aloud” Task*

Task Scores		Instruction Scores	
Construction of Knowledge	3	Higher Order Thinking	3
Elaborated Communication	4	Deep Knowledge	4
Connection to Students’ Lives	1	Substantive Conversation	3
		Connectedness to the Real World	1
Total Task:	8	Total Instruction:	11

### **Generalizations**

The difficulties associated with implementing instruction consistent with the authentic pedagogy model are well documented (Onosko, 1991; Rossi, 1995; Saye & Brush, 2004). The authentic pedagogy scores from this study suggest that some teachers

were more successful than others in overcoming these challenges. In this section, I identify and discuss some trends that were apparent among the teachers in each of the authentic pedagogy categories. Due to the inherent complexity associated with teaching and the classroom environment, these assertions should be viewed as very tentative.

Roy and Andy's practice varied the most from the authentic pedagogy model. Andy seemed to equate good teaching with getting students to pass the graduation exam. When asked to submit tasks that demonstrated students thinking at a high level, he selected a research project, powerpoint presentation, and a chapter worksheet. The nature of these tasks suggested that he assigned greater significance to tasks that required the mastery of larger quantities of factual information. Higher order thinking and reasoning goals were noticeably absent. Like Andy, Roy's tasks also tended to reinforce basic knowledge (i.e. teach a lesson) with the possible exception of the illustrated timeline task.

It is possible that the pedagogical beliefs of these teachers conflicted with elements of the AIW model. Their decision to focus primarily on transmitting factual knowledge to students could stem from any number of factors (influence of the high stakes test, perceived ability of students, belief that basics must be learned before advanced work can be considered, time involved for inquiry lessons, heritage based view of history, etc.). It was difficult to determine exactly what motivated their instructional decision-making since I was unable to schedule an interview with Andy and Roy's interview was relatively brief.

My personal sense from observing these lessons was that these teachers were engaged in defensive teaching (McNeil, 1986). Teachers who engage in defensive teaching limit the knowledge they make accessible to students in order to efficiently

cover information and maintain classroom control. Roy's defensive teaching probably stemmed from his inexperience and difficulties with classroom management. He didn't appear willing to "rock the boat" very often by pushing students towards more challenging work (Sizer, 1984). I believe that students become more proficient in completing authentic tasks and respond more favorably to them with routine exposure and coaching from the teacher. I believe this is one reason why Lee (from the moderate category) seemed to have more success in implementing the same illustrated timeline activity Roy used with his class. Lee's students had encountered similar activities in the past and had a better sense of what was expected.

Andy's motives for engaging in defensive teaching might have been similar to Roy's since his classes tended to be large and fairly diverse. However, the defensive teaching technique that was most noticeable during observations of his lessons was simplification of knowledge (McNeil, 1986). Andy appeared to possess relatively strong content knowledge, but he simplified topics in order to more efficiently move through the curriculum. Lesson material was covered with relatively little debate or discussion. I didn't get the sense that Andy was trying to engage students in the examination of conflicting interpretations of the past. Andy's goal, from what I could tell, was to strictly stick to the requirements identified in the course of study.

Jason's teaching was on the borderline between minimal and limited authentic pedagogy. Some of his tasks were challenging (i.e. rewriting the Declaration of Independence), but they either weren't very interesting to students or they didn't include enough support to help most students be successful. While Jason seemed open to allowing more student inquiry, his dominant instructional approach was likely more

traditional. Student comments during the observed lessons, sometimes very revealing, supported this conclusion.

In summary, students who experienced minimal levels of authentic pedagogy were rarely pushed to think and reason at levels beyond basic recall and comprehension. The tasks submitted by teachers in the limited authentic pedagogy category (Amy and Phillip) were not always that different. However, there was evidence that these teachers had internalized certain elements of the authentic pedagogy model and were more receptive to an inquiry-based instructional approach. Their tasks were more likely to include higher order thinking elements and elaborated forms of communication. Students had the opportunity to create documentaries, participate in debates, and analyze paintings; tasks that were very different from those in the minimal category.

While students were afforded these opportunities, the teachers in the limited category sometimes struggled to provide the support necessary for quality work. The lessons didn't reach their full potential for a number of reasons. In Amy's case, most of her activities came from pre-packaged curriculums (i.e. *History Alive*). These activities promoted active learning and at least some higher order thinking. However, the teachers at the moderate level likely achieved higher implementation scores, in part, because they were involved in the creation (or at least modification) of their tasks. They were able to take their lessons to a deeper level as a result of the "sweat equity" involved in the creation process.

The lack of a debriefing also caused some lessons to not reach their full potential. There may have been an assumption that if thought provoking ideas were presented during the course of a lesson, then students could synthesize them on their own. During

one of Amy's lessons in particular, students debated the views of various philosophers regarding the best form of government. However, they were not afforded the opportunity to step out of character and discuss their own perceptions of the ideas being expressed. They were expected to complete a demanding follow-up essay without any closure to the lesson. Successful inquiry teachers view the debriefing of a lesson as the big "pay off"; the time when students are pushed to make connections and think at a higher level.

Attempts to elicit higher order thinking were also sometimes undermined by the actions of the teacher. For example, Phillip moved quickly through the Manifest Destiny material in order to be able to spend more time on the Civil War. This sent a message to students about the relative importance of the challenging task. Even in instances where the task was fully implemented, the teachers in this category sometimes sent an unintentional message through their assessment practices by emphasizing the easier to grade lower order aspects of the activity.

The teachers in the moderate authentic pedagogy category probably had greater success achieving higher scores because their vision of powerful social studies instruction most closely aligned with the authentic pedagogy model. Lee and Ryan were both very articulate in expressing their goals for social studies instruction. They clearly identified the need to create competent citizens as the overarching purpose for social studies instruction. Their curricular decision-making was driven by goals related to this purpose. They wanted students to not only build content knowledge, but the capacity to think and make decisions. In addition, they wanted students to connect history instruction with contemporary issues.

The experience for students in classrooms which featured moderate levels of authentic pedagogy was different from those mentioned previously in several important ways. First, students in Ryan and Lee's classroom were more likely to be challenged with meaningful historical problems that required higher order thinking (Czar Nicholas Think Aloud, Berlin Crisis Think Aloud, Industrial Revolution Editorial, etc.). The higher order elements of the activities usually took precedence over everything else. The students appeared to be accustomed to these types of challenges and aware that the challenging aspects of the assignment were going to be evaluated.

The student experience was also different in terms of the support they received during the learning process. Ryan and Lee had a keen understanding of the cognitive demands being placed on their students. They helped students manage these demands in a number of ways. They provided thorough instructions and made their expectations clear by going over examples and non-examples of quality student work. Students received hard scaffolds to help them with the thinking tasks embedded in the activities. These were not the worksheets or handouts associated with pre-packaged curriculum materials. These were frequently designed by the teacher and placed at strategic points in the lesson. When the hard scaffolding wasn't enough, Ryan and Lee were able to effectively diagnose student difficulties and provide timely scaffolding without diluting the overall challenge of the task. Finally, the students were more likely to participate in a debriefing after challenging activities. The debriefings were another form of scaffolding. They were instrumental in helping the class develop some shared understandings about the lesson prior to follow-on individual assignments.

Ryan and Lee clearly understood the challenges and opportunities associated with inquiry based instruction. They possessed key dispositions often associated with successful inquiry teachers. They were open to constructivist ideas about learning and the nature of historical knowledge. They also possessed the internal drive and intelligence needed to successfully negotiate the cognitive challenges associated with inquiry based instruction. The scores they achieved in this study were likely the result of sustained professional development through the pursuit of advanced degrees and active involvement in a professional learning community focused on advancing inquiry based teaching.

This chapter has described the range of instructional experiences students encountered in their social studies classes. It seems clear that the experiences of students in classrooms with higher levels of authentic pedagogy were quite different from those in the lowest category. The next chapter will present research findings related to the effects of different levels of authentic pedagogy on the acquisition of basic content knowledge. It will also discuss the results of the higher order editorial assessment.

## CHAPTER FIVE: STUDENT LEARNING OUTCOMES

In the previous chapter, I organized the sample of teachers in this study into three categories (minimal, limited, and moderate) based on their authentic pedagogy scores. In doing so, I provided examples of the tasks and instruction students received at these levels to highlight different ways teachers conceptualized intellectual challenge. The purpose of this chapter is to analyze the effects of authentic pedagogy on student learning. I begin with a review of the sample used in this study. This is followed by the results of the study presented in order by research question.

### **Description of the Sample**

This study included eight social studies teachers. Four of the teachers taught 9<sup>th</sup> grade World History at the junior high level. The remaining teachers taught 10<sup>th</sup> grade social studies. Every 10<sup>th</sup> grade social studies teacher at the high school in 2008 was a participant in the study. Information about the teacher sample (i.e. demographics, experience) was presented in the previous chapter and is reproduced in Table 25 for easy reference.

A broad range of student level data (anonymous to the researcher) was collected as part of this study. Table 26 presents descriptive statistics associated with the student database. Additional information about the student database is provided in Appendix T.



Table 25

*Teacher Profiles*

	Roy	Andy	Jason	Amy	Phillip	Lauren	Ryan	Lee
AP Score	9.6	10.9	11.6	12.9	13.3	18	20.9	21.2
Age	26-35	36-35	36-45	46-55	26-35	46-55	26-35	36-45
Ethnicity	White	White	White	White	White	White	White	White
Experience	4	11	14	22	6	15+	11	12
Grade Taught	9	10	10	9	10	9	10	9

Table 26

*Descriptive Statistics for Student Sample*

	2008 (N=351) Percent	2009 (N=454) Percent
Gender		
Male	49.3	50.0
No Data	5.1	2.4
Ethnicity		
White	63	55.7
African American	24.8	26.7
Asian	6	5.5
Hispanic	1.1	1.5
Not Reported/No Data	5.1	10.5
SES – based on lunch status		
Paid	74.1	69.8
Reduced	4	4.8
Free	16.8	15.2
No Data	5.1	10.1
Special Education		
Yes	6	4.8
No/No Data	94	95.2
English Proficiency		
Limited	3.4	3.1
Proficient	91.5	85.9
No Data	5.1	11
New to System (arrived 2006-2009)		
Yes	17.7	15.4
No	76.6	74.4
No Data	5.7	10

## **Results of Inferential Analyses**

**Research Question II.** Do students that have been taught by teachers demonstrating higher levels of authentic pedagogy score higher on the Alabama High School Graduation Exam (AHSGE) than students taught by teachers with lower levels of authentic pedagogy? Null Hypothesis: The level of authentic pedagogy a student receives in their tenth grade social studies course does not have a statistically significant effect on students' graduation exam scores.

The first step in analyzing this question was to conduct a content analysis of the graduation exam. The content analysis, using item specifications released to the public, confirmed that the test was a measure of lower order knowledge and therefore appropriately being used in this study. Results of the analysis are explained in greater detail in Appendix S.

I used two different statistical approaches to address this research question. First I used multiple regression with students as the unit of analysis. This had some significant limitations since the students were only associated with four possible teachers. In an attempt to gain more meaningful results, I also ran ANOVA tests comparing specific classes with varying levels of authentic pedagogy. The results of both of these approaches are described in sequence in this section.

**Regression.** In conducting the multiple regression analysis, several initial models were produced to ascertain whether any predictor variables overlapped in explaining student performance. After identifying and eliminating the areas of overlap (see Appendix U for further discussion of this process), the final model included 427 students who took regular 10<sup>th</sup> grade United States history over the course of the two years

covered by the study. The results of the regression analysis are provided in Table 27. The overall model was able to account for 44% of the variance in the social studies graduation exam scores. Demographic variables had the most influence on achievement (26%). When tenth grade social studies averages were added, another 15% was explained. With all of these variables controlled, authentic pedagogy was able to contribute an additional 3%.

Table 27

*Sequential Multiple Regression Analyses Predicting Impact of Authentic Pedagogy on Graduation Exam Results*

Model	Variable	R Square/ Change	Beta	Semi-partial
1. Demographics		.261***		
	Gender		.235***	.229***
	Ethnicity		.189***	.139***
	LEP		-.047	-.047
	Special Ed		-.141***	-.138***
	SES		.068	.050
2. Achievement		.149***		
	10 <sup>th</sup> Average		.444***	.380***
3. Authentic Pedagogy		.027***		
	Task Instruction		-.151***	-.134***
			.163***	.144***
OVERALL MODEL		.437***		

*Note.* N=427; \*p<.05, \*\*p<.01, \*\*\*p<.001

The best predictor of scores on the graduation exam was a student's tenth grade average in social studies. This was followed by a student's gender and then the level of

authentic instruction they received. Authentic instruction had a positive effect on student graduation exam scores while authentic tasks had a negative influence. In both instances the relationship was significant, although the negative influence of authentic tasks was not very strong when compared to the ethnicity and special education variables.

*ANOVA.* In order to better gauge the effects of authentic pedagogy, I also tried using the class as the unit of analysis. A class level analysis required selecting two similar classes for comparison from teachers who utilized different levels of authentic pedagogy. I first paired a class from the minimal authentic pedagogy category (Andy) with one from the limited authentic pedagogy category (Phillip). I conducted statistical tests to ensure significant variances didn't exist between the classes on key variables likely to influence achievement (i.e. demographics, social studies grades, etc.). These tests are described in Appendix V.

The one-way ANOVA comparing achievement on the graduation exam between the minimal and limited authentic pedagogy classes indicated that the minimal authentic pedagogy class performed significantly better,  $F(1, 44) = 9.516$ ,  $MSE = 2591.5$ ,  $p = .004$ ,  $\eta^2 = 0.18$ . Table 28 provides additional details regarding the performance of the two classes.

Table 28

*One-way ANOVA Comparing Graduation Exam Scores for Minimal & Limited Classes*

<b>Class</b>	<b>Mean SS AHSGE</b>	<b>SD</b>	<b>Range</b>	<b>Difference</b>
Limited AP	512.27	46.809	491.52 to 533.03	F=9.516
Minimal AP	558.62	54.380	535.66 to 581.59	

This finding should be considered with some caution. While a number of variables were controlled, it is still possible that uncontrolled variables played a role in contributing to the difference in outcomes (i.e. teacher variables such as experience in the classroom, etc.). Also, the two teachers were fairly close on the authentic pedagogy scale (10.9 compared to 13.3). A subsequent analysis comparing Andy's class with another period of Philip's yielded similar results that were not statistically significant.

I also compared Andy's minimal authentic pedagogy class with a class taught by the highest scoring tenth grade teacher (Ryan). Both classes were regular U.S. History courses, although Ryan taught a class that met on alternate days for the entire school year. Ryan and Andy's classes were similar in terms of gender, SES, and prior social studies achievement (see Appendix V). The main area of concern was race. The difference between classes was significant with Andy's class having the larger number of African Americans. In order to make a more valid comparison, I focused my analysis on white students only.

The results of this analysis revealed a difference in graduation exam scores that differed very slightly in favor of the low authentic pedagogy class, but the results were likely due to chance,  $F(1, 29) = .000$ ,  $MSE = 3033.055$ ,  $p = .986$ . The moderate authentic pedagogy class had the advantage of a year round schedule and their social studies grades were slightly higher. This might suggest that there should have been a greater difference in the mean scores in favor of the moderate AP class. On the other hand, authentic instruction focuses primarily on learning outcomes that are not measured by the graduation exam. Since the cut score on the graduation exam was a 509 and the mean

score of the moderate class was 558, white students in Ryan’s class were clearly not put at a disadvantage on this test.

Table 29

*One-way ANOVA Comparing Graduation Exam Scores for Minimal & Moderate Classes*

<b>Class</b>	<b>Mean SS AHSGE</b>	<b>SD</b>	<b>Range</b>	<b>Difference</b>
Moderate AP	558.45	51.855	534.18 to 582.72	F=.000
Minimal AP	558.82	60.720	518.03 to 599.61	

Once again, caution is in order in interpreting the results of this analysis. A variety of additional factors that were not controlled in this analysis could explain the difference in performance between these two classes. The next step after analyzing the impact of authentic pedagogy on lower order outcomes was to determine its effect on another type of assessment designed to measure more advanced thinking processes.

**Research Question III.** What is the impact of authentic pedagogy on student performance on an assessment that requires them to apply knowledge from a previous unit to a challenging new task?

Null Hypothesis: The level of authentic pedagogy a student receives in their tenth grade social studies course does not have a statistically significant effect on their ability to apply knowledge from a previous unit to a challenging new task.

In order to address this research question I created a writing task that required students to construct an editorial based on an authentic historical problem. The tasks and rubrics are discussed in greater detail in chapter three. The regular and AP editorials

were analyzed separately. I will begin by discussing the results of the regular U.S. History assessment.

I began my analysis by organizing the database of students who took the Manifest Destiny higher order editorial into three groups based on the level of authentic pedagogy they experienced (1=minimal, 2=limited, 3=moderate). The minimal group included all the students who took history from Andy or Jason. The limited group included all the students who took history from Phillip. The remaining students in the moderate group took Ryan's history classes. The unit of analysis was students within the three large groups, not specific classes. After establishing the three groups, I wanted to see if they differed significantly on specific demographic characteristics (race, gender, SES). In taking this step I was attempting to control for factors, other than authentic pedagogy, that might influence student performance. Appendix W provides more information regarding the statistical tests I used to establish the comparability of the groups.

The final step was to run a factorial MANOVA using the rubric sub-categories for Part I of the editorial as dependent variables: position, context, persuasiveness, low level dialectical reasoning, and quality of final position (see rubric in Appendix N). The independent variables were the level of authentic pedagogy students experienced (as represented by the three teacher groups) and race. Race was included as an independent variable because I was not able to establish that the three groups had similar black/white ratios (Hotelling's Trace  $p=.039$ ). The results associated with the descriptive statistics are presented here first before examining the outcome of the factorial MANOVA.

The total score from Part I of the editorial had a possible range of 0 to 14. The distribution of scores is provided in Table 30. The total score is derived from several

scoring categories on the rubric. Some of these, like position and historical context, can be a little deceptive because students can score relatively well if they simply follow directions. The persuasiveness and low-level dialectical reasoning scales provide a clearer window into what students were able to do on this assessment. It is for this reason that I decided to highlight these categories. The table indicates that nearly half of the students scored between 0 and 3 on Part I. No students reached the top end scores of 11-14. Analysis of the persuasive and dialectical categories revealed a similar pattern. Most students did not reach the upper end of these scales. Student editorials, for the most part, were not very persuasive. Students struggled to provide elaborated arguments that were backed by historical evidence and/or examples. The low-level dialectical reasoning scale measured the extent to which students were able to identify and explain the opposing viewpoint from the one they were arguing. Most students (53%) either did not include opposing arguments or did not provide enough information to clearly demonstrate they understood an opposing point of view. Students who provided opposing views often immediately refuted them in the same paragraph without ever clearly laying out a well developed opposing perspective. The lack of a strong third paragraph made it difficult for students to achieve a high score on the final paragraph where more advanced dialectical reasoning was measured.

Part II of the editorial assessed whether students saw any connection between Manifest Destiny and contemporary U.S. policies. The highest scores (level 2) were reserved for students who explicitly mentioned Manifest Destiny in their response and tied it to their explanation of America's mission (or lack of a mission) in the world today. A small number of the responses made this type of connection (4.5%).



Table 30

*Distribution of Manifest Destiny Editorial Scores for Part I and Select Rubric Sub-Categories*

Total Score - Part I		Persuasiveness Score – (Part I)		Low-Level Dialectical Reasoning Score – (Part I)	
Score	Percent	Score	Percent	Score	Percent
0	7.1	0	27.7	0	52.9
1-3	40.6	1	32.9	1	33.5
4-6	34.8	2	18.7	2	12.9
7-9	14.8	3	19.4	3	.6
10	2.6	4	1.3		
11-14	0	5	0		

*Note.* N=155

The results of the factorial MANOVA are described in Table 31. Note that while race was included as a variable in this analysis, it did not reach significance in terms of its impact on student performance (Hotelling's Trace  $p=.107$ ). A statistically significant difference was found between the level of authentic pedagogy a student experienced (minimal, limited, or moderate) and their academic performance on the regular U.S. History higher order editorial task. I ran Bonferroni post-hoc tests to determine more specifically how the level of authentic pedagogy influenced achievement. The results indicated that the moderate authentic pedagogy group performed significantly better than the minimal group ( $p=.001$ ) and the limited group ( $p=.018$ ) on the context scale. In addition, the moderate authentic pedagogy group performed significantly better than the limited group ( $p=.001$ ) on the persuasiveness scale. However, there was not a significant effect when comparing the moderate group with the minimal group on this component of

the rubric. The moderate group did perform at a higher level, but this could have been due to chance. These results should be viewed with some caution. While inter-rater reliability figures are available for the advanced placement editorial, they are not for this editorial. It is possible that another rater could come to different conclusions. Two additional social studies graduate students did read numerous Manifest Destiny editorials and assisted with the process of refining the rubric.

Table 31

*Factorial Manova Comparing the Performance of Authentic Pedagogy Groups on the Manifest Destiny Editorial*

Variable	Minimal Mean (SD)	Limited Mean (SD)	Moderate Mean (SD)	F
Position	.71 (.457)	.67 (.480)	.71 (.457)	.078
Context	.46 (.703)	.48 (.643)	.92 (.651)	6.059*
Persuasiveness	1.29 (1.175)	.78 (.847)	1.69 (1.071)	4.164*
Low-level Dialectical Reasoning	.58 (.792)	.30 (.465)	.81 (.754)	2.886
Quality of Final Position	.49 (.653)	.44 (.577)	.81 (.754)	2.382

*Note.* An overall multivariate comparison resulted in a Hotelling's Trace of .163 ( $p=.019$ )  
\* $p<.05$ .

In reading through the Manifest Destiny editorials some general trends were apparent. Quite a few students didn't have a firm grasp of the historical context associated with the conflict with Mexico leading up to the decision point in 1846. Introductory paragraphs reflected confusion over a number of factual details. Students demonstrated misconceptions over who won the Battle of the Alamo, Mexico's relationship with Texas (i.e. the fact that Texas was a part of Mexico before the conflict with the U.S.), and whether the Americans who migrated to Texas were invited or not. There was also the tendency to overlook the border dispute which most directly

precipitated the Mexican-American War. Some students also used the term “Manifest Destiny” in strange and awkward ways suggesting a lack of in-depth understanding of what it meant (i.e. the U.S. will use the Manifest Destiny on you). Students who did use the term correctly weren’t necessarily able to provide much more than a basic definition. I expected to see more instances of students making historical connections to reinforce the notion that America had a special God-given destiny (i.e. a City upon a Hill reference from earlier in the U.S. History 10 course).

The writing prompt for the Manifest Destiny task also seemed to confuse many students. They were inclined to argue for or against the Mexican-American War without discussing the concept of Manifest Destiny. This was a substantial problem because I really expected students to focus their response around their understanding of Manifest Destiny and how this term was being used at the time. The students who wrote a limited response without discussing Manifest Destiny received a maximum of two points on the persuasiveness scale for Part I.

Students used a number of arguments to support their position and write a persuasive editorial. Those who believed American actions related to the Mexican-American War were justified often claimed that Mexico was at fault for refusing to meet with the American representative, John Slidell or refusing to sell land to the United States. Some students argued that Mexico was really the aggressor by attacking U.S. troops in the disputed territory. When students integrated Manifest Destiny into their response (required for a higher score), they argued to a greater or lesser extent that U.S. actions were for the greater good because America would bring advances (i.e. democracy) to a land that couldn’t seem to stabilize its government following its

independence from Spain (an argument from the Boston Times source document). Those who argued against America's actions towards Mexico frequently used the argument by Albert Gallatin that America was simply acting out of greed. This was often paired with the idea that Manifest Destiny was being used as a "cover" to obfuscate America's real intentions. Of course the argument that God didn't really support Manifest Destiny was commonly used as well. There was little evidence of students critically examining and weighing the arguments contained in the documents. For example, students used the statement "before this unfortunate war, [America] always acted with justice... The use of military force was always in self-defense..." without ever seeming to question its veracity. Several students used a phrase from the Boston Times article justifying American actions because they would better the lives of the "great mass of the people, who have, for a period of 300 years been the slaves of an overbearing foreign race." While this was a legitimate argument of the time, it is interesting to note that only one student referred to America's own problem with slavery.

*German Unification Editorials.* I used a similar procedure to evaluate the advanced placement editorials. Instead of having three authentic pedagogy groups, I only had two since I didn't have an advanced placement teacher score in the minimal range. Once again, I attempted to control for any differences between the groups on factors that might impact achievement on the editorial assessment. The limited and moderate authentic pedagogy groups were not significantly different in terms of race, gender, or socio-economic status (see Appendix W).

The descriptive statistics associated with this assessment are reported in Table 32. Nearly half of the students either did not provide a clear position statement in their

editorial or their statement focused on whether Germany should unify instead of *how* unification should be accomplished (i.e. include Austria?). Most of the students provided at least some background context in setting up the editorial (>70%). Most students scored at the two or three level on the persuasiveness scale indicating editorials that were adequate at best in convincing the reader. Adequate editorials generally included two or more persuasive reasons that were described without much elaboration or support. In terms of lower level dialectical reasoning, most students (52.2%) did not provide enough information in the third paragraph to indicate they had a solid understanding of an opposing perspective. The final position standard of the rubric evaluated the ability of students to provide a persuasive conclusion that also included higher level dialectical reasoning: genuine consideration of opposing viewpoints. Most students (58.9%) scored at the adequate level (1) indicating that they included a basic conclusion that restated some key points. A small group of students (7.8%) scored at a higher level (2) by providing a more elaborate conclusion that added to the overall persuasiveness of the editorial. No students scored a three which would have required evidence of advanced dialectical reasoning. The scores for Part I of the editorial had a possible range of zero to fourteen. No students scored over ten. Most students scored in the four to six range (47.9%). A fairly sizeable group (24%) scored from seven to nine. Only 1.1% achieved a score of ten.

Part II of the assessment was evaluated on the basis of decision-making and persuasiveness. The majority of the students scored at the two level out of a possible five points. When combining parts I and II, the range of possible scores was from one to nineteen. No students earned a score over thirteen.

Table 32

*Distribution of German Unification Editorial Scores*

Position Sub-Scale		Context Sub-Scale		Persuasiveness Sub-scale		Dialectical Reasoning Sub-scale		Part I		Part II		Total Score	
Score	Percent	Score	Percent	Score	Percent	Score	Percent	Score	Percent	Score	Percent	Score	Percent
0	44.5	0	26.7	0	6.7	0	52.2	1	6.7%	0	21.1	1	4.4%
1	55.6	1	48.9	1	20	1	28.9	2	7.8%	1	14.4	2	4.4%
		2	24.4	2	46.7	2	15.6	3	12.2%	2	31.1	3	6.7%
				3	25.6	3	3.3	4	16.7%	3	24.4	4	7.8%
				4	1.1			5	15.6%	4	8.9	5	5.6%
				5	0			6	15.6%	5	0	6	16.7%
								7	14.4%			7	12.2%
								8	6.7%			8	7.8%
								9	3.3%			9	20%
								10	1.1%			10	6.7%
								11-14	0%			11	3.3%
												12	2.2%
												13	2.2%
												14-19	0

*Note.* N=90

As with the Manifest Destiny editorials, I ran a factorial MANOVA to determine if any differences existed in the performance of students from the authentic pedagogy groups (limited & moderate) on the higher order editorial. In this analysis, I examined the rubric sub-categories from Part I. In virtually all instances, the moderate group achieved higher mean scores than the limited group. However, this result did not reach significance in any scoring category. Table 33 provides additional information associated with this analysis.

Table 33

*Manova Comparing the Performance of Authentic Pedagogy Groups on the Advanced Placement German Unification Editorial*

Variable	Limited Mean (SD)	Moderate Mean (SD)	F
Position	.50 (.508)	.59 (.496)	.673
Context	1.00 (.696)	.96 (.738)	.052
Persuasiveness	1.76 (.890)	2.05 (.862)	2.320
Low-level Dialectical Reasoning	.56 (.786)	.79 (.889)	1.502
Quality of Final Position	.62 (.551)	.82 (.606)	2.556

*Note.* An overall multivariate comparison resulted in a Hotelling's Trace of .062 (p=.400).

Returning to the null hypothesis for this research question, the results suggest higher levels of authentic pedagogy had a positive impact on student performance on the higher order writing task. However, the difference in performance for AP students could be due to chance. The null hypothesis was rejected for the regular history classes and retained for the advanced placement groups.

As with the Manifest Destiny editorial, the biggest issue I encountered was off-topic responses. Instead of addressing the question of whether a unified German state should include *all* Germans, the students often argued the merits of unification itself. It was common for students to discuss how a larger unified state would increase Germany's military might and prestige among the nations of Europe. In this context, they would often incorporate Mohl's statement (from the primary source document) of how a Reich with 70 million people could not be challenged. The practical problems associated with forming a unified country encompassing German speaking territories in Austria were generally not discussed or if they were, only superficially. Students also discussed the economic benefits of being united and perhaps how a country formed by people with similar customs, language, and beliefs could function more smoothly. Some of the off-topic papers were fairly well written, but they did not exceed a two (out of 5) on the persuasiveness scale because they did not really address the issue of nationalism and the extent to which it should serve as the basis for German unification.

Among the students who did address the appropriate question, most editorials did not exceed a level three, "adequate" score for persuasiveness (see Appendix X for examples). When scoring persuasiveness I evaluated the entire editorial. However, the main portion of the editorial dedicated to providing supporting argumentation was paragraph two. Two of the better supporting paragraphs are provided in Figure 5. These paragraphs demonstrate some of the arguments constructed by students and how the source documents were integrated into the editorials. The excerpt on the left provides an argument that unification only serves the interests of Prussia and Bismarck. It suggests that unification would not follow the liberal constitutional course favored by the people.



Despite a factual error, this editorial is noteworthy in that only one other student made this sort of argument. Also, the student included relevant information from the textbook that was not in any of the source documents. Later in the editorial (not in the second paragraph) the student also makes a political argument. The language used in the “opposing” editorial on the left was not particularly eloquent or clear. Also, the student could have taken a more decisive stance on the focus question. As a result an adequate (3) persuasiveness score was assigned. The excerpt on the right makes military and economic arguments for unification. While it did not contain lengthy arguments, it did contain some original elaboration particularly in the area that discusses ports in the Mediterranean and coal mines in the Saar. The overall editorial earned an elaborated (4) persuasiveness score.

The use of the primary source documents by the students often amounted to pulling brief passages to supplement an argument or to represent an opposing view. The stronger students integrated the quotes into their argument with a short explanation. In a number of instances, however, students simply inserted a quote in the paragraph and moved on. This lack of elaboration and some of the clearly inaccurate statements made in conjunction with quotes suggested that the documents were fairly difficult for many of the students to understand. The students particularly had a hard time with the following statement: “The highest and most fundamental idea in the political life of a state must be the internal satisfaction of peoples through their institutions, their right to self-determination.” This sentence could have been used to frame an argument against basing a state solely on nationalism, but most students missed the point Bebel (the author of the speech) was trying to make. Instead, some students interpreted the statement to mean

that the goal of a state is to ensure people are happy and unification would accomplish this objective.

Opposing Unification	Supporting Unification of All Germans
<p>In the 1860s Prussia wanted to increase military power. In order to do that, the parliament had to appropriate the budget to finance it. The parliament disapproved such action, which shows that most people of Prussia didn't want it. But, Bismarck spent government money on military strength, ignoring the parliament. It shows that the decision was by few authority, not the people. And in 1848, there was an assembly called Frankfurt. It was a meeting of German states to propose a plan for the unification of Germany. The plan was to have a constitutional monarch, that is to be more liberal than before. Prussian king refused, saying that unification should be obtained by blood. If Prussia actually wanted unification for the German states, they would have agreed at the assembly. But since they wanted a strong power centered in Prussia, they decided to obtain unification through war. The two evidence shows that Prussia was solely decided and the decision was by a few authorities. Thus, it doesn't need to be supported, not only by German people, but also by foreign nations.</p>	<p>The unification of Germanic peoples should be endorsed because of the potential military they would gain from such a unification. If Germany were to unify they would, according to document 2 and the words of Mohl, have a Reich of seventy million people. This new Reich would be able to stand up to even Russia with her sixty-six million, and France with her thirty-six million. Having such a superior force will make Germany powerful and that means unification will help them a lot. Another consequence that I believe makes unification good for Germany is the economic consequence. With the unification of such a vast amount of land it stands to reason that they would also gain many different forms of economic stimulus. They would gain ports in the Mediterranean in Austria-Hungary and coal mines in Saar. This alone will much help the German economy and make unification well worth it.</p>

*Figure 5.* Examples of Supporting Arguments Provided for Part I of the German Unification Editorial

As with the Manifest Destiny editorials, I occasionally got the sense that students were appropriating arguments from the primary source documents without really examining their underlying logic. For example, students cited the same Bebel speech in

their editorial to demonstrate that nationalism would lead to war and that a unified state based on nationalism would require Germany to cede territory away (i.e. Slavic-speaking areas). It is true that this point was made in the primary source document. However, I thought at least one student would question this claim. Would those seeking to unify Germany really give away territory in the name of nationalism? It seems doubtful that any state would voluntarily do this. One student did bring up the scenario of ethnic cleansing to develop a pure German state. This argument is probably not very authentic for a citizen living in 1870, but it does suggest the student was thinking about the implications of the arguments being made in the source document.

**Research Question IV.** Does the ability to apply knowledge on the graduation exam improve with repeated exposure (multiple courses) to classroom experiences that require students to perform challenging intellectual tasks? Null Hypothesis: Repeated exposure to authentic classroom experiences that require students to perform challenging intellectual tasks has no impact on student performance on the Alabama High School Graduation Exam.

A one-way ANOVA was used to test whether repeated exposure to moderate levels of authentic pedagogy resulted in higher scores on the Alabama High School Graduation Exam. I created a variable representing the total number of social studies courses each student had at the moderate authentic pedagogy level. This resulted in three possibilities: a student could have 0, 1, or 2 social studies classes at the moderate level in their 9<sup>th</sup> and 10<sup>th</sup> grade years. Students who had more than one social studies teacher in any particular year were filtered out of the analysis. The resulting ANOVA included 328

students with no exposure to moderate pedagogy, 292 with one class, and 58 students in the group who experienced two moderate authentic pedagogy courses. Table 34 provides a breakdown of the results (first row of data). Repeated exposure to courses featuring moderate authentic pedagogy was found to have a significant effect on student achievement on the graduation exam ( $p < .001$ ). Although the ANOVA showed that the means were significantly different, the effect size was very small ( $\eta^2 = .04$ ). The Eta squared was just .04 when .10 is needed for a small effect according to Cohen (Cohen, 1992, p. 157).

Hochberg’s GT2 post-hoc comparisons of the three groups indicated that the group with two moderate authentic pedagogy classes had significantly higher scores on the graduation exam than students with one. This group also performed significantly better than students who did not have any experiences at the moderate level ( $p < .001$ ). The same relationship was found when comparing students who experienced one moderate authentic pedagogy course with those who didn’t experience any at the moderate level. The students with one course performed significantly better,  $p = .004$ .

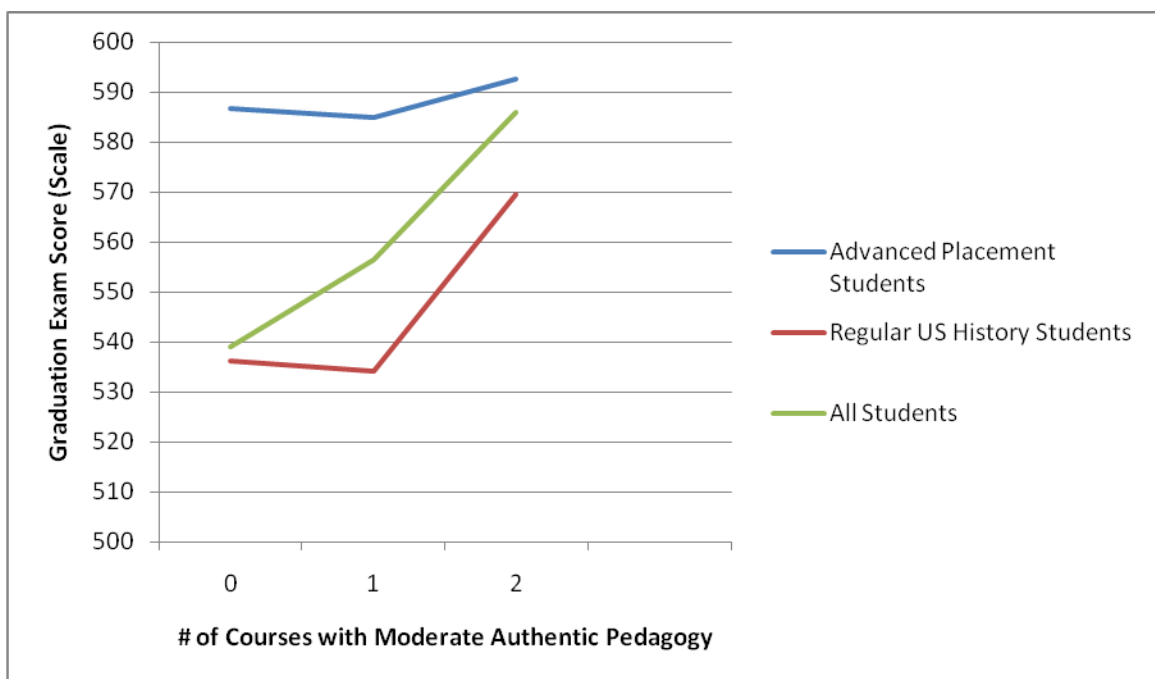
Table 34

*Analysis of the Impact of Courses Featuring Moderate Authentic Pedagogy on Graduation Exam Results*

	No moderate AP classes Mean (SD)	One moderate AP class Mean (SD)	Two moderate AP classes Mean (SD)	F
Advanced Placement Students included	539.08 (70.257)	556.38 (65.686)	585.91 (47.924)	14.13***
Advanced Placement Students excluded	536.16 (70.263)	534.12 (64.369)	569.53 (51.762)	2.121

Note. \*\*\* $p < .001$

Each group in this analysis included regular and advanced placement history students. However, a larger percentage of advanced placement students were in the ANOVA group that had the most repeated exposure to authentic pedagogy. It is possible that the results of the analysis were influenced by an advanced placement effect instead of just authentic pedagogy. This seems likely since advanced placement students tended to do better on the graduation exam than students in the regular U.S. history course. In order to more precisely examine the research question, I ran another one-way ANOVA that excluded the advanced placement students (second row of data in Table 34). Only 17 students were in the group that experienced two social studies courses at the moderate authentic pedagogy level. The test did not indicate a statistically significant difference between the three groups. Figure 6 graphically displays the results of these two tests as well as an ANOVA that included only advanced placement students.



*Figure 6.* Effect of repeated exposure to moderate authentic pedagogy on student achievement. The green “all students” line was the only one to reach statistical significance. Results were significant at the .01 level from 0 to 1 and 1 to 2.

Finally, I applied the same predictor variable to a sequential regression model similar to the one I used to address research question two. The prior moderate variable was entered by itself in model three. In this analysis, I included advanced placement students since the previous ANOVA indicated that only a very small sample of non-AP students had multiple classes with moderate authentic pedagogy. Multiple social studies courses with moderate authentic pedagogy had a slight positive impact on student achievement on the graduation exam. The results are displayed in Table 35. Based on the results of the ANOVA models and the regression analysis, the null hypothesis was rejected.

Table 35

*Sequential Multiple Regression Analyses Predicting Impact of Repeated Exposure to Moderate Authentic Pedagogy on Graduation Exam Results*

Model	Variable	R Square/ Change	Beta	Semi- partial
1. Demographics		.263***		
	Gender		.194***	.190***
	Ethnicity		.169***	.129***
	LEP		-.102***	-.102***
	Special Ed		-.106***	-.104***
	SES		.097**	.073**
2. Achievement		.147***		
	10 <sup>th</sup> Average		.441***	.389***
3. Authentic Pedagogy		.016***		
	Prior Moderate		.131***	.128***
OVERALL MODEL		.426***		

*Note.* \*p<.05, \*\*p<.01, \*\*\*p<.001

**Research Question V.** To what extent does authentic pedagogy bring different achievement benefits to students of different social and academic backgrounds? Null Hypothesis: Authentic Pedagogy will result in statistically significant differences in achievement on the graduation exam for students from different social and academic backgrounds.

I used several bivariate correlation tests to address this question. In order to maintain consistency with previous analyses, I excluded students who had more than one

social studies teacher in the tenth grade and students in advanced placement courses. I also decided to analyze this question using three variables: authentic pedagogy, authentic tasks, and authentic instruction. I examined the influence of authentic tasks and instruction independently to gain a more nuanced understanding of how the components of authentic pedagogy impact the performance of various subgroups of students.

Table 36 depicts the bivariate correlation examining the relation between authentic pedagogy and student performance. The results suggest that authentic pedagogy positively impacted graduation exam performance for all subgroups. The correlation for white, male students reached statistical significance at the .05 level. The effect size in both cases (gender and race) was small. A significant difference in achievement benefit did not exist based on socio-economic background or prior academic achievement. The results of the bivariate correlations for authentic pedagogy led me to accept the null hypothesis for gender and race, and to reject the null hypothesis for SES status and prior achievement.

The analysis of authentic tasks by themselves yielded a different outcome. Based on this limited sample of teachers, authentic tasks were often negatively correlated with student performance on the graduation exam. There was not a statistically significant performance benefit associated with authentic tasks based on a student's gender, SES, or prior social studies achievement. The performance of African American students was negatively associated with authentic tasks at the .05 level, but the effect size was small. White students experienced a positive effect, but they could have just as easily experienced a negative one since the outcome was not statistically significant.



Table 36

*Bivariate Correlations Examining Relation between Authentic Pedagogy and Achievement by Subgroups*

	AP Mean	Pearson Correlation	Effect Size
<b>Gender</b>			
Female (N=214)	14.170	.107	0.01144
Male (N=280)	13.388	.139*	0.01932
<b>Ethnicity</b>			
White (N=281)	13.932	.127*	0.01612
African-American (N=168)	13.407	.051	0.00260
<b>SES</b>			
Paid Lunch (N=334)	13.840	.095	0.00902
Free/Reduced Lunch (N=135)	13.565	.136	0.01849
<b>Social Studies Achievement</b>			
A/B Student (N=305)	14.025	.081	0.00656
C/D/F Student (N=174)	13.264	.010	.0001

*Note.* \* $p < .05$

Table 37

*Bivariate Correlations Examining Relation between Authentic Tasks and Achievement by Subgroups*

	Task Mean	Pearson Correlation	Effect Size
<b>Gender</b>			
Female (N=214)	6.675	-.110	.0121
Male (N=280)	6.508	.025	.000625
<b>Ethnicity</b>			
White (N=281)	6.611	.005	.000025
African-American (N=168)	6.549	-.180*	0.0324
<b>SES</b>			
Paid Lunch (N=334)	6.601	-.065	0.00422
Free/Reduced Lunch (N=135)	6.569	-.058	0.00336
<b>Social Studies Achievement</b>			
A/B Student (N=305)	6.634	-.043	0.00184
C/D/F Student (N=174)	6.483	-.158	0.02496

*Note.* \* $p < .05$

Finally, I examined the performance benefits of authentic instruction as it related to achievement among the same subgroups of students. Authentic instruction was positively correlated with student achievement on the graduation exam, regardless of a student's demographic profile or social studies achievement record. The more authentic instruction students received, the better they performed on the graduation exam. The results achieved statistical significance for both genders, white students, and more advantaged students (based on paid lunch). They approached significance for low SES students ( $p=.054$ ). Authentic instruction was positively associated with performance for students with both low and high social studies averages, but the results could have been due to chance. The comparison for each variable did not reveal any drastic differences among subgroups in how authentic instruction influenced performance.

Table 38

*Bivariate Correlations Examining Relation between Authentic Instruction and Achievement by Subgroups*

	Instruction Mean	Pearson Correlation	Effect Size
Gender			
Female (N=214)	7.495	.160*	0.0256
Male (N=280)	6.880	.157*	0.024649
Ethnicity			
White (N=281)	7.320	.149*	0.022201
African-American (168)	6.858	.131	0.017161
SES			
Paid Lunch (N=334)	7.239	.134*	0.017956
Free/Reduced Lunch (N=135)	6.996	.185	0.034225
Social Studies Achievement			
A/B Student (N=305)	7.390	.109	0.011881
C/D/F Student (N=174)	6.782	.072	0.005184

*Note.* \* $p<.05$

## Summary

In this chapter I've provided the results of analyses related to four research questions. My objective was to better understand how authentic pedagogy influences student learning in history classrooms. The findings suggest that authentic pedagogy has a small, but positive impact on student performance on the Alabama High School Graduation Exam. However, other factors such as grades in social studies and gender are stronger predictors. Classroom level comparisons suggest that students who receive higher levels of authentic pedagogy are not put at a significant disadvantage on the AHSGE. It is important to note that they also did not experience the sort of performance benefit that would be consistent with the outcomes reported in Newman's 2001 study of authentic pedagogy and standardized tests. In Newman's study, students who received high quality assignments were likely to outperform their peers on the standardized tests by significant margins (i.e. an achievement benefit of 32 points on the IGAP reading section). The results of my analysis should be viewed with caution. This study dealt with a very limited sample of teachers (N=4). The spread of these teachers along the authentic pedagogy continuum was also limited with no teacher reaching the substantial level and only one teacher in the moderate category.

This study also had an equity component. I was interested in whether any performance benefits associated with authentic pedagogy would be equally distributed among the students. Authentic pedagogy was found to have a positive impact on student performance for all social and prior achievement groups. White, male students experienced a greater achievement benefit than African-Americans and females. I also examined the sub-components of authentic pedagogy: authentic tasks and authentic

instruction. When the analysis focused solely on authentic tasks the results, in most instances, suggested a negative influence on student performance on the graduation exam when students experienced tasks that scored higher in authenticity. Male and female students were impacted about the same, as were students from different socio-economic and academic backgrounds. The negative impact on African-American students reached statistical significance when compared to the very small positive impact of authentic tasks on the performance of white students. The impact of higher levels of authentic instruction on student performance, on the other hand, was positive in most instances. The results of bivariate correlations revealed no significant difference in performance benefit based on gender or prior academic achievement. African American students were positively impacted by higher levels of authentic instruction, but not to the same extent as whites. The same was true for students from lower socio-economic backgrounds based on free or reduced lunch when compared to students that paid for their lunch.

Another area I chose to investigate was the impact of exposure to multiple classes with higher levels of authentic pedagogy. Would students who experienced two courses of moderate authentic pedagogy perform better on the graduation exam than students who only receive one or none at all? The answer to this question was “yes” based on my analysis. However, the results of this question are a little more complicated due to the nature of the student sample. Many of the students who were in the group who received multiple classes of moderate authentic pedagogy were also advanced placement students. It was difficult to separate the effect of authentic pedagogy versus a possible advanced placement effect.

Finally, I looked at the impact of authentic pedagogy on higher order learning outcomes. I chose to focus my statistical analysis for both tasks on the rubric sub-categories associated with Part I because they were the most reliable in terms of scoring and they offered a targeted look at specific higher order skills (i.e. persuasiveness). The cumulative scoring categories were less meaningful because they could be influenced to a greater degree by procedural aspects of the scoring rubric. When looking at the editorials written by the regular U.S. History students, the students who received moderate authentic pedagogy wrote more persuasive editorials than classmates who received limited authentic pedagogy. The same was true when comparing the moderate group with students who received minimal authentic pedagogy, but the results could have been simply due to chance. There was not a statistically significant difference among classes of advanced placement students on the German Unification task. The students in the limited authentic pedagogy group generally did not perform as well as the students in the moderate authentic pedagogy group on the rubric sub-categories associated with Part I, but this could have been due to chance. The next chapter will provide a more extended discussion of these findings.

## CHAPTER SIX: SUMMARY, LIMITATIONS, & IMPLICATIONS

This study investigated the impact of authentic instruction on student learning in social studies classrooms. As discussed in the literature review, there is evidence to suggest that today's high-stakes tests serve as a disincentive for those who want to provide more in-depth learning experiences for their students. Teachers need reassurance that they are not hurting their students' chances on high stakes tests when they pursue more ambitious, and often time consuming, inquiry-oriented activities. Work by Newmann in other subject areas provides evidence that authentic pedagogy can enable students to achieve positive results on basic skills tests while also producing complex intellectual learning outcomes. This study was an effort to extend this line of inquiry in the social studies. I wanted to more fully understand the types of learning outcomes students demonstrate when they receive higher levels of authentic pedagogy in their history classes.

In order to operationalize and analyze the instruction students experienced, I used Newmann's Authentic Intellectual Work (AIW) model. This framework places greater value on teaching that encourages higher-order thinking, in-depth knowledge, substantive communication, and real world application - characteristics commonly associated with inquiry-based instruction. Participating teachers in the study were categorized using AIW rubrics and placed on a continuum according to the level of authentic pedagogy they provided to their students.

Once teachers were categorized, I created a database that included students from the participating teacher's classes. Each student record included demographic information, prior achievement data, and social studies graduation exam results. I conducted statistical analyses using the database to determine how students that experienced varying levels of authentic pedagogy performed on measures of lower and higher order knowledge.

In previous chapters I've discussed the theoretical basis for this study, its methodology, and findings. This chapter offers a more extended discussion of some of the major findings. It places the results within the context of those from similar studies. Alternative explanations for the results of the study are provided as well as suggestions for further research.

## **Summary**

This study included five research questions. The first research question was: *To what extent do teachers utilize authentic pedagogy and how much variation exists within the sample of teachers in this study?* I concluded that high levels of authentic pedagogy were not very prevalent in the study schools. The range of possible authentic pedagogy scores (7-30) was broken down into four categories to reflect a continuum from minimal use of authentic pedagogy to substantial. No teachers in this sample provided substantial authentic pedagogy. However, a good deal of variation still existed among study teachers with the lowest score being 9.6 and the highest 21.2. Three teachers were in the moderate authentic pedagogy category, two in the limited, and three in the minimal category. The average score in this sample was a 14.8 which was below the mean of the scale of 18.5.

The second research question focused on lower order learning outcomes associated with authentic pedagogy. The question asked: *Do students that have been taught by teachers demonstrating higher levels of authentic pedagogy score higher on the Alabama High School Graduation Exam (AHSGE) than students taught by teachers with lower levels of authentic pedagogy?* I analyzed this question in several ways. The most precise analysis involved comparing a class that experienced minimal authentic pedagogy with one that received moderate authentic pedagogy. The class that experienced minimal authentic pedagogy outperformed the moderate authentic pedagogy class on the graduation exam, but the results were not statistically significant. The results of this analysis led me to conclude that authentic pedagogy did not cause students to perform at a higher level on this test of basic historical knowledge. However, it did not appear to hurt students' chances either. With only four teachers in this analysis, these results should be viewed as very tentative.

I also examined the same question using students as the level of analysis instead of intact classes. The results of this broader analysis suggested that authentic pedagogy played a small positive role in explaining student performance. When the elements of authentic pedagogy were analyzed independently, authentic tasks were found to have a negative impact on student performance, while authentic instruction had a positive impact. The results in both cases were statistically significant.

The third research question examined the impact of varying levels of authentic pedagogy on higher order learning outcomes. The question asked: *What is the impact of authentic pedagogy on student performance on an assessment that requires them to apply knowledge from a previous unit to a challenging new task?* The “challenging new task”



was an editorial assignment that required students to take and defend a stance on a historical problem. Regular students completed an editorial focused on Manifest Destiny while the advanced placement students did a similar assignment on German Unification. Most students struggled on the editorial assignment, regardless of their assignment as advanced or regular.

Analysis of student editorials revealed a statistically significant difference in the three group's performance on the Manifest Destiny editorial. Students in the moderate authentic pedagogy group were able to write editorials that contained better introductory paragraphs with more historical context than those in the minimal and limited groups. They also wrote more persuasive editorials, although the result only reached statistical significance when the moderate group was compared with the limited authentic pedagogy group. In analyzing the advanced placement editorials, students were organized into two groups: limited and moderate authentic pedagogy. The students in the moderate authentic pedagogy group had higher mean scores on most of the rubric categories (i.e. persuasiveness, dialectical reasoning, etc.) that comprised Part I of the assessment task, but the differences were not statistically significant. In general, both the general and advanced students were able to use the documents provided to construct basic arguments for or against the question under consideration.

The fourth research question asked whether the ability to apply knowledge on the graduation exam improved with repeated exposure (multiple courses) to classroom experiences that required students to perform challenging intellectual tasks. Students performed better on the graduation exam when they had two classes of moderate pedagogy as compared to having just one experience or none at all. When all students

were included in the analysis the results were statistically significant. However, most of the students who experienced two classes at the moderate level were in the Advanced Placement course in the tenth grade. It is difficult to know whether the higher scores on the graduation exam were because students experienced two moderate level courses or simply because the students were more academically advanced. When the AP students were eliminated from the analysis, the results were not statistically significant.

Finally I asked: *To what extent does authentic pedagogy bring different achievement benefits to students of different social and academic backgrounds?*

I analyzed the achievement benefits associated with authentic pedagogy, authentic tasks and authentic instruction as part of this question. Authentic pedagogy was positively correlated with achievement on the graduation exam regardless of a student's prior academic ability or demographic group. The achievement benefits were equitably distributed for students based on SES status and academic ability (prior social studies grades). However, white, male students experienced a greater achievement benefit than African-Americans and women.

When authentic tasks were analyzed independently as a sub-component of authentic pedagogy, the resulting output indicated a negative correlation for the SES and prior academic achievement variables. In other words, higher scoring tasks were associated with lower performance on the graduation exam for students with these demographic and achievement characteristics. The correlation did not reach statistical significance for either variable (free/reduced vs. paid lunch or A/B students vs. C,D,F students). When analyzing the influence of authentic tasks based on gender, the correlation with student performance was positive for males, but negative for females.

However, the difference was not statistically significant. Finally, the correlation between authentic tasks and student performance on the graduation exam for ethnicity was positive for whites and negative for African-Americans. The difference was significant at the .05 level. This suggested that the use of authentic tasks had a greater impact on African-Americans than whites, and was associated with lower achievement for this group of students. However, the effect size for this analysis was small (0.03).

Authentic instruction, the other sub-component of authentic pedagogy, was associated with improved performance for all groups. The more authentic instruction students received, the better they performed on the graduation exam regardless of their demographic profile or prior achievement, although the correlation did not always reach statistical significance. The lack of statistical significance was particularly evident for African-American students and those from lower socio-economic backgrounds.

## **Discussion and Alternative Explanations**

**AIW and Lower Order Achievement Outcomes.** One of the main aspects of this study was determining the impact of authentic pedagogy on the Alabama High School Graduation Exam (AHSGE). The results of my study are consistent with other AIW studies that suggest authentic pedagogy does not hurt student performance on standardized tests (D'Agostino, 1996; Lee, Smith, & Croninger, 1997). However, they do not support the findings from the study that most directly addressed this relationship (Newmann, Bryk, & Nagaoka, 2001). Newmann's 2001 study indicated that students (grades 3, 6, 8) who received higher quality authentic tasks performed *at higher levels* on basic skills tests in reading, writing, and math. These results were explained in terms of vocabulary acquisition and motivation to learn. Newmann argued that AIW's ability to

promote these benefits essentially offset any limitations imposed by reduced coverage of testable material.

Why did AIW not have the same impact on student retention of lower order knowledge in this study? One possibility is simply the fact that I had such a small sample of teachers at the grade level the graduation exam was administered (N=4). Larger samples may have yielded results more similar to those of past AIW studies. The sample was also less than ideal since it did not include any teachers at the substantial authentic pedagogy level. Perhaps higher levels of authentic pedagogy among the teacher sample are needed to achieve the outcomes found in Newmann's study. Another explanation could have to do with the outcome measure itself. The graduation exam covers a significant period of U.S. history (Beginnings to WWII) and measures retention of specific information. It is possible that Newmann's theory regarding vocabulary acquisition and motivation doesn't hold true for high school achievement tests of this nature.

This study differed from most AIW studies in that it examined both tasks and instruction in determining the authentic pedagogy scores. This research design enabled me to examine the impact of tasks and instruction independently and in conjunction with the overall authentic pedagogy score. The negative impact of authentic tasks on student performance was initially puzzling to me, but upon further reflection makes sense. It is possible that the teachers adopted more challenging tasks, perhaps as a result of professional development or to impress me, without altering their usual instruction to any great extent. A similar theory was suggested in some of the Gates Foundation research that used the authentic intellectual work model to analyze school reform (AIW/SRI,

2007). If this is true, the students might have been unable, due to inadequate preparation, to fully take advantage of the learning opportunity represented by the authentic tasks. Not only would they struggle to achieve the higher order objectives associated with the task, but they might also grow frustrated or confused to the point where the lower order objectives were compromised. There are examples from the study of moderate scoring tasks coupled with minimal levels of authentic instruction (see Appendix O – Jason and Phillip). It seems logical that challenging tasks, by themselves, would not really provide a big boost in student achievement. It is very difficult for teachers who do not routinely challenge their students to produce positive outcomes with challenging tasks “on demand” or immediately. Students may need opportunities to build up to the challenging tasks. When I observed certain tasks being implemented, it was apparent that the students were experiencing something out of the norm. For instance, when observing Jason’s lesson that required students to rewrite the Declaration of Independence in contemporary language, it seemed clear that students were being confronted with a task that was much more challenging than usual. A student asked Jason whether they were going “back to hard.” Jason replied by saying that it [the lesson] could be challenging, or simple and enjoyable depending on the day. My sense was that Jason offered occasional instructional challenges, but this was not a consistent focus of the class. I got a much different impression when watching the teachers in moderate authentic pedagogy classes. It is very difficult to successfully plan, scaffold, and implement authentic tasks. The moderate authentic pedagogy teachers might have scored higher on authentic instruction because it was something they more routinely did with their students. It is possible that the routine helped hone the skills of the teacher while also conditioning students to react

more favorably when challenged. My hypothesis is that authentic tasks would have a more positive impact on student performance if they were a more consistent focus of the teacher. Another aspect of this study was determining if exposure to multiple courses at higher levels of authentic pedagogy resulted in improved learning outcomes. Analysis revealed that students who had multiple years of authentic pedagogy at the moderate level generally had higher graduation exam scores than their peers in classes with lower levels. However, the finding was not significant when advanced placement students were removed from the equation. The results of this study are not nearly as strong as those identified in Klentschy's research. Klentschy and his associates compared the performance of elementary science students on two standardized assessments based on whether they experienced a constructivist based, hands-on science program or a more traditional curriculum. The students in the constructivist oriented program outperformed the students who did not experience the program and their performance improved steadily as they experienced more years of the program (Klentschy, Garrison, & Amaral, 2001). The disparity between the results in my study and those reported by Klentschy et. al. could simply be a result of the small sample size in my study or the fact that the grade level and discipline were different. It was also very difficult to separate and clearly define the impact of authentic pedagogy over multiple years from the impact of advanced placement courses. Perhaps a different study design would have yielded better results.

**AIW and Higher Order Achievement Outcomes.** As mentioned at the beginning of this chapter, the higher order research findings were based on an analysis of a writing task completed by the students. The task provided to the advanced placement students focused on the issue of German unification and the regular U.S. History students

completed an editorial focused on Manifest Destiny and the Mexican-American War. My hypothesis in designing the editorial assessments was that students who routinely experienced instruction requiring them to critically examine ideas and formulate arguments would be able to develop a well reasoned, persuasive editorial based on a problem they hadn't encountered previously. They would be able to recognize holes in logic or the implications of arguments being made in the source documents even if they were not particularly well versed in all of the details associated with German unification or Manifest Destiny. Overall, there was not a great deal of evidence of this type of thoughtful reflection to support my hypothesis. However, this could be a result of a number of extraneous factors that have little to do with the overall ability of the students to engage in this type of thinking.

The results could be attributed to the novelty of the task. The AP students, in particular, were accustomed to document-based questions. A scaffolded-essay of this type might have appeared foreign to them. My own experience with students of this age suggests that the students in either group (regular or advanced) were probably capable of providing better responses with more guidance. In attempting to standardize the assignment instructions provided by the teacher, I was left "out of the loop" and thus unable to answer questions or intervene to address misconceptions students might have had about the assignment (i.e. clarifying what the question was asking). In retrospect, I should have attempted to gain permission to administer the assignment myself.

Teachers were also told to provide incentives for the students so they would try hard on the assessment. At a minimum I wanted the assignment to be graded so students would have some stake in their performance. I have no way of knowing if all of the

teachers followed through on this request. It is possible that some classes were more motivated to give their best than others.

Another complication is that I did not observe the instruction students received on this topic. Although I made a careful effort to consult with the teachers on the broad topic of each editorial, it is possible that they emphasized different aspects of Manifest Destiny and German Unification in their classes. Some students, in covering Manifest Destiny for example, may have received a blow by blow account of the battles associated with Texas' independence and the Mexican-American War with little discussion of the motives of the participants. I also don't know how much instructional time was dedicated to each topic. If the students were really uncomfortable with the topic, they might not have performed up to their true potential. Finally, although every effort was made to make the editorials as engaging and relevant as possible it is possible that this type of task did not appeal to some students and this could have also influenced their effort.

Even with these considerations, the results of the higher order task were still revealing. It would be interesting to compare student scores on the Manifest Destiny editorial with their subsection scores from the graduation exam that dealt with the same topic. It is possible that many of the students might be able to correctly answer multiple choice questions, but the editorials suggest that most students do not adequately understand this time period or the concept of Manifest Destiny. While the editorial task is more challenging to score, it certainly provides a better window into student misconceptions of history. Another conclusion I draw from reading the editorials, both AP and regular, is that students probably need more opportunities to engage in tasks that



develop higher order skills such as persuasive argumentation and dialectical reasoning. It is important for students to be able to “think like a judge,” evaluate a problem from multiple angles, and develop defensible solutions. This has important implications for the future of democracy.

### **Limitations**

Despite the incredible generosity of the school system and the willingness of the social studies faculty to invite me into their classrooms, some limitations still existed in this study. First, I lacked the resources to evaluate student work. The use of all three AIW rubrics makes it easier to form judgments regarding the level of intellectual challenge in the pedagogy students experienced as part of their coursework. However, other rigorous studies of this nature have been conducted without student work. My design was particularly strong in that it linked tasks with classroom observations. This enabled me to more accurately ascertain the teacher’s intent and to see how the teacher’s instructional approach might either add to or detract from the intellectual challenge associated with a particular task.

Another limitation was the presence of interns in some of the study classrooms. Interns were required to teach a minimum of twenty days with the full load of classes. Some exceeded this amount based on guidance from their cooperating teacher. This study did not include an evaluation of intern instruction to determine its level of authenticity. It was therefore difficult to determine the impact of their instruction on student learning. However, I had student data for each of the cooperating teachers from semesters where no intern was involved in instruction. Also, one assumes that cooperating teachers supervised interns closely to ensure the standards they’ve

established for their course were met. A teacher is likely to intervene if an intern is not teaching the things he/she believes are important. However, in most cases one would not expect a novice teacher to provide the same level of authentic instruction as a skilled veteran.

My association with the study participants was another limitation. I knew most of the teachers through professional development seminars and contacts associated with my assistantship (i.e. supervision of interns) at Auburn University. The potential certainly existed for bias in rating. Ideally, the second rater used for inter-rater reliability (IRR) would have no relationship with the teachers involved in the study. This was the case when it came to achieving inter-rater reliability for tasks in this study since they were often evaluated by other SSIRC researchers. However, I did not have the resources to train an outside researcher or provide compensation for travel to the observations. Instead, my advisor, Dr. Saye served as the second observer. Dr. Saye and I came to a shared understanding of how to apply the rubrics to instruction as a result of the training associated with the SSIRC project. I believe that this understanding significantly reduced the likelihood that a teacher would systematically be rated lower due to personal bias.

My association with the study teachers had another effect on this study. It seemed at times that some teachers might have been trying to “game the system” by turning in tasks they thought I’d like. This is understandable, but an ideal scenario would involve teachers forming an independent judgment of what to submit based on a professional sense of what constitutes instructional quality. While the study teachers were not familiar with authentic intellectual work per se, they did know of my association with curriculum development projects that adhered to a problem-based historical inquiry model. Since

this model closely relates to AIW, some teachers were able to guess that I wanted tasks that challenged students to apply historical knowledge and think critically. To the extent that teachers had inquiry tasks on hand, this might have inflated some task scores. However, it is hard to fake the standards associated with the instruction rubric. Some teachers might have been better served (in terms of their authentic pedagogy score) by submitting tasks that better fit their comfort level to execute.

I was also limited by the number of blocks that I could reasonably observe. I tried to not only vary the three observations across the course of a semester, but also the blocks that I observed. If a teacher taught AP and regular courses, I tried to see lessons associated with each. However, this was not always possible. An assumption of this research is that a teacher does not vary his/her instruction significantly from block to block (or from year to year). More extensive interviews (pre-post) would have possibly helped to determine if this assumption was valid for each participant in the study.

It is difficult to make wide ranging generalizations from this study since it included a very limited sample of teachers and used outcome measures not found in other states. The scoring of the higher order editorials was also very challenging given that the range in student performance was not always great. Subtle distinctions and judgments had to sometimes be made to arrive at scores. The rubric could probably still be improved to enhance its reliability.

### **Implications and Areas for Further Study**

The results of this study raise a number of questions and areas for further research. First of all, it is perhaps troubling that authentic pedagogy is not more

prevalent in history classrooms. Most of the teachers in this study were classified as providing minimal or limited authentic pedagogy. This is disconcerting since the study schools represented one of the best possible areas in the state to look for inquiry-based instruction (in terms of resources, reputation, professional development, etc.). The results of this study are consistent with the broader SSIRC study that also documented relatively low levels of authentic pedagogy of most school settings (Social Studies Inquiry Research Collaborative, 2011). The implication of this from a policy standpoint is that considerable time, effort, and resources are likely needed to cultivate meaningful changes in teacher practice. Various professional development initiatives have been implemented in the past to improve the capacity of teachers to provide authentic pedagogy. More studies, like those conducted by Avery in the late 1990s, are needed to determine the most effective ways to help teachers not only conceptualize challenging tasks, but also provide students with the support they need to be successful (Avery, Kouneski, & Odendahl, 2001; Avery & Palmer, 1999). The AIW scoring rubrics are a powerful tool that should be used more widely by districts and schools to improve instruction.

Policy-makers and school officials may be unwilling to promote authentic pedagogy without greater evidence of its impact on student learning. My study was too limited in scale to determine conclusively how authentic pedagogy influences student performance on standardized history tests. I can tentatively conclude that it doesn't hurt student achievement. Further research is needed to confirm the relationship between authentic pedagogy and standardized tests in social studies. The research needs to include a larger sample of teachers and a greater variety of standardized assessments. It is possible that certain types of standardized social studies assessments are more likely to

reveal performance benefits from higher levels of authentic pedagogy. Would the results of this study be different if the lower order measure was an end-of-course test instead of an assessment that included eleventh grade material the students hadn't covered yet? Would it be different if the U.S. History National Assessment of Educational Progress (NAEP) was used as the achievement measure or a test from another subject area like Economics?

The strongest claim made for adopting authentic pedagogy has been its potential for securing higher order learning outcomes. Policy-makers would probably be willing to make a stronger commitment to this model if it could be tied to gains in 21<sup>st</sup> century skills. The evidence of improved learning is not as strong in social studies as compared to other subject areas. Researchers often don't know what students in the "control" classes are capable of doing, because they aren't given the opportunity to complete authentic tasks in their more traditional classroom settings. There are not many studies, like this one, that compare students who experience varying degrees of authentic pedagogy on a common task that requires higher order thinking. Therefore, it is hard to say with certainty that inquiry-based instruction, as defined by the authentic pedagogy model, is more effective than instruction completely dominated by lecture or some other approach.

The results of the higher order portion of this study are just as tentative as those that dealt with the graduation exam. Analysis of the editorial assessments did not reveal a large performance benefit from higher levels of authentic pedagogy (the effect size was very small even if the results were statistically significant). Students in all classes in the sample generally struggled with the task, whether they were regular or advanced. This

seems to support the view of critics who contend that students lack the foundational knowledge or developmental characteristics needed to complete higher level challenges. However, this finding can be misleading. What is perhaps not fully captured, because I was not able to use audio and video equipment during classroom observations, is the difference from a qualitative standpoint in what I observed in the study classrooms. When asked to complete tasks that required significant higher order thinking, many students in the moderate authentic pedagogy classes were able to rise to the occasion (a fact consistently noted in Newmann's own research). Students in the minimal or limited classes either were not afforded the same opportunities or they did not respond as favorably for a number of reasons. The assessment instruments that I created for this study were not as effective as I would have liked in measuring the range of higher order outcomes associated with authentic pedagogy. The results might have been different had the assessment format allowed for the soft scaffolding and peer support students normally receive in a class setting.

More research is needed to develop common higher order assessments that can be reliably scored under similar study conditions to evaluate the impact of different levels of authentic pedagogy on student performance. They need to move beyond paper and pencil tasks to include scored discussions and other alternative projects likely to get at the sort of outcomes authentic pedagogy is designed to elicit. A big problem is the overloaded testing schedule at most schools. Researchers should continue to focus on states, like Washington, that already use classroom based assessments (CBAs) as part of their accountability program. Ideally, partnerships between teachers, researchers, and other stakeholders would result in the development of innovative and authentic CBAs. Then

researchers could analyze student performance over a larger geographic area in relation to the instruction they received. The work on “rich tasks” in Queensland, Australia provides a good model for this sort of endeavor.

Another area related to this study that deserves follow-up related to this study is the compounding effect of authentic pedagogy. This study really doesn't provide a definitive answer to the question of whether student achievement improves with multiple courses at the moderate authentic pedagogy level. The performance trend was positive, but it could have been due to chance or other factors discussed in the last chapter. Longitudinal studies, with a larger sample of teachers and students, are needed to further investigate this question. This represents a significant research challenge since the difficulty associated with achieving top scores on the AIW rubrics is well documented. It will likely take some effort to locate a suitable setting where a substantial sample of students experiences a succession of courses at the moderate authentic pedagogy level or higher.

Several additional areas are also worthy of additional study. The Gates Foundation studies separated their analysis of the tasks provided by teachers to allow for two variables: rigor and relevance. They examined whether a task's rigor or relevance played a bigger role in influencing student achievement (AIR/SRI, 2007). The rigor variable was similar to Newmann's construction of knowledge standard. The relevance variable measured the extent to which students could have a voice and influence in what they were being asked to do, whether the task connected to the real world, and if it involved something adults might be plausibly asked to do (AIR/SRI, 2006). While the Gate's researchers found a task's rigor to be more directly correlated to quality study

work in English and math, this might be different in a study focused on history learning outcomes. I believe social studies researchers should more closely examine the relevance (“connectedness to students’ lives) portion of the authentic intellectual work model. The connectedness scores in my study were noticeably low. Most teachers did not attempt to make explicit connections between the historical topic they were studying and contemporary issues or events. To what extent would student achievement have been greater in this study if modifications were made to improve the scores associated with just this one particular standard? What is the unique impact of making a task more relevant to a student’s life when it comes to history?

Finally, future studies would probably benefit from collecting tasks, observing instruction, and analyzing student work. While this can be challenging to accomplish, it would likely provide the best picture of what students experienced in their social studies class and it would allow for the most precise classification of teachers along the authentic pedagogy continuum. The classification of teachers is important since it forms the basis for analyzing student learning outcomes and really determining the impact of authentic pedagogy.

## **Conclusion**

This study aimed to better understand the learning outcomes associated with authentic pedagogy. Numerous studies dealing with the authentic intellectual work construct have suggested that teachers who assign more challenging work to their students receive products of higher quality when compared to teachers who don’t offer their students the same types of opportunities. These studies, however, have often dealt with subjects other than social studies. Very few of them investigate how authentic



pedagogy influences student performance on standardized tests. This study attempted to address a need in the field by examining the impact of authentic intellectual work on student achievement in history. The results of the study suggest that authentic pedagogy does have a positive influence on student learning, but not to the extent demonstrated by most of Newmann's studies. However, there is room for cautious optimism. This study suggests that student performance on high-stakes tests is not compromised when teachers utilize more in-depth, inquiry oriented instructional approaches. The positive impact of authentic pedagogy may grow for students who experience multiple classes that reach at least the moderate level as defined in this study. The effects of authentic pedagogy are equitably distributed among most significant sub-groups of students within schools (i.e. gender, race, SES). These findings will be revisited as data from this study are analyzed in conjunction with the larger Social Studies Inquiry Research Collaborative project. Hopefully, as the pilot study for this effort, my research will contribute in a small way towards providing teachers with useful information that will help them to improve their practice and better serve students.

## References

- Achieve Inc. (2004). *Do graduation exams measure up? A closer look at state high school exit exams*. Washington, D.C.: Achieve, Inc.
- Aikin, W. M. (1942). *The story of the Eight-Year Study*. New York: Harper.
- AIR/SRI. (2004). Exploring assignments, student work, and teacher feedback in reforming high schools: 2002-03 data from Washington State. Retrieved Mar. 8, 2008, from [http://www.air.org/expertise/index/?fa=viewContent&content\\_id=300](http://www.air.org/expertise/index/?fa=viewContent&content_id=300)
- AIR/SRI. (2006). Evaluation of the Bill & Melinda Gates Foundation's high school grants initiative: 2001-2005 final report. Retrieved March 8, 2008, from <http://www.gatesfoundation.org/learning/Documents/Year4EvaluationAIRSRI.pdf>
- AIR/SRI. (2007). Changes in rigor, relevance, and student learning in redesigned high schools: An evaluation for the Bill and Melinda Gates Foundation. Retrieved March 8, 2008, from [http://www.air.org/reports-products/index.cfm?fa=viewContent&content\\_id=295](http://www.air.org/reports-products/index.cfm?fa=viewContent&content_id=295)

Alabama Department of Education. (2009a). Chief State School Officer's Report for Alabama High School Graduation Exam. Retrieved August 14, 2009, from <http://www.alsde.edu/Accountability/2009Reports/CSSO/CSSOAHSGE.2009.pdf?lstSchoolYear=7&lstReport=2009Reports%2FCSSO%2FCSSOAHSGE.2009.pdf>

Alabama Department of Education. (2009b). Process used to determine cut scores for the Alabama High School Graduation Exam. Retrieved Sept. 20, 2009, from <http://www.alsde.edu/text/sections/documents.asp?section=91&sort=1&footer=sections>

Amosa, W., Ladwig, J., Griffiths, T., & Gore, J. (2007). *Equity effects of quality teaching: Closing the gap*. Paper presented at the Australian Association for Research in Education Conference, Fremantle.

Armstrong, N. (1970). *The effect of two instructional inquiry strategies on critical thinking and achievement in eighth-grade social studies*. (Unpublished doctoral dissertation): Indiana University, Bloomington, IN.

Avery, P. G. (1999). Authentic instruction and assessment. *Social Education*, 65(6), 368-373.

Avery, P. G., Bird, K., Johnstone, S., Sullivan, J. L., & Thalhammer, K. (1992). Exploring political tolerance with adolescents. *Theory and Research in Social Education*, 20(4), 386-420.

- Avery, P. G., Kouneski, N. P., & Odendahl, T. (2001). Authentic pedagogy seminars: Renewing our commitment to teaching and learning. *The Social Studies* (May/June), 97-101.
- Avery, P. G., & Palmer, E. (1999). *Professional development for authentic pedagogy in the social studies: An evaluation*. Minneapolis: The Center for Applied Research in Educational Improvement.
- Bain, R. (2000). Into the breach: Using research and theory to shape history instruction. In P. Stearns, P. Seixas & S. Wineburg (Eds.), *Knowing, teaching, and learning history: National and international perspectives* (pp. 331-353). New York, NY: University Press.
- Baldi, S., Perie, M., Skidmore, D., Greenberg, E., & Hahn, C. (2001). *What democracy means to ninth graders: U.S. results from the International IEA civic education study*. Washington, D.C.: U.S. Department of Education, National Center for Education Statistics.
- Barratt, T. K. (1964). *A comparison of effects upon selected areas of pupil learning of two methods of teaching United States history to eleventh grade students*. (Unpublished doctoral dissertation).
- Bartlett, F. C. (1932). *Remembering*. Cambridge, MA: Harvard University Press.
- Barton, K. (1997). "I just kinda know": Elementary students' ideas about historical evidence. *Theory and Research in Social Education*, 25(4), 407-430.

- Barton, K. C. (2008). Research on students' ideas about history. In L. S. Levstik & C. A. Tyson (Eds.), *Handbook of Research in Social Studies Education* (pp. 239-258). New York: Routledge.
- Barton, K. C., & Levstik, L. S. (2004). *Teaching History for the Common Good*. Mahwah: Lawrence Erlbaum Associates.
- Bayles, E. E. (1956). Experiments with reflective teaching. In *Kansas studies in education* (pp. 32). Lawrence, KA: University of Kansas Publications.
- Bennett, W. J. (1992). *The de-valuing of America: The fight for our culture and our children*. New York, NY: Simon and Schuster.
- Berlak, H., Newmann, F. M., Adams, E., Archbald, D. A., Burgess, T., Raven, J., et al. (1992). *Toward a new science of educational testing and assessment*. Albany, NY: SUNY Press.
- Boote, D. N., & Beile, P. (2005). Scholars before researchers: On the centrality of the dissertation literature review in research preparation. *Educational Researcher*, 34(6), 3-15.
- Bransford, J. D., Brown, A.L., Cocking, R.R. (Ed.). (2000). *How people learn: Brain, mind, experience, and school*. Washington, D.C.: National Academy Press.
- Braun, H. I. (2005). *Using student progress to evaluate teachers: A primer on value-added models*. Princeton, NJ: Educational Testing Service.

- Brooks, J. G., & Brooks, M. G. (1993). *In search of understanding: The case for constructivist classrooms*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher*, 18(1), 32-42.
- Bruner, J. S. (1960). *The process of education*. Cambridge: Harvard University Press.
- Byungro, S. (1991). *The comparative effects of problem-solving instruction and conventional expository instruction on students' acquisition, retention, and structuring of knowledge in high school social studies*. (Unpublished doctoral dissertation). University of Georgia, Athens, GA.
- Cheney, L. (1994). The end of history. *Wall Street Journal*.
- Chenoweth, R. W. (1953). *The development of certain habits of reflective thinking*. (Unpublished doctoral dissertation). University of Illinois, Urbana, IL.
- Cizek, G. J. (1991a). Effusion confusion: A re-joinder to Wiggins. *Phi Delta Kappan*, 73, 150-153.
- Cizek, G. J. (1991b). Innovation or enervation? Performance assessment in perspective. *Phi Delta Kappan*, 72(9), 695-699.
- Cognition and Technology Group at Vanderbilt. (1990). Anchored instruction and its relationship to situated cognition. *Educational Researcher*, 19(5), 2-10.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112, 155-159.

- Conley, D. (2003). *Mixed messages: What state high school tests communicate about student readiness for college*. Eugene, OR: Center for Educational Policy Research, University of Oregon.
- Cornbleth, C. (1985). Critical thinking and cognitive processes. In W. B. Stanley (Ed.), *Review of research in social studies education: 1976-1983. Bulletin No. 75* (pp. 11-63). Washington, D.C.: National Council for the Social Studies.
- Cousins, J. E. (1962). *The development of reflective thinking in an eighth grade social studies class*. (Unpublished doctoral dissertation). Indiana University, Bloomington, IN.
- Cox, B. C. (1961). *A description and appraisal of a reflective method of teaching United States history*. (Unpublished doctoral dissertation). Indiana University, Bloomington, IN.
- Cronin, J., Dahlin, M., Adkins, D., & Kingsbury, G. G. (2007). *The proficiency illusion*. Washington, D.C.: Thomas B. Fordham Institute.
- Curtis, C. K., & Shaver, J. P. (1980). Slow learners and the study of contemporary problems. *Social Education*, 44, 302-309.
- D'Agostino, J. V. (1996). Authentic instruction and academic achievement in compensatory education classrooms. *Studies in Educational Evaluation*, 22(2), 139-155.

- Daugherty, R. (2004). *Getting high school graduation test policies right in SREB states*. Atlanta, GA: Southern Regional Education Board.
- De La Paz, S. (2005). Effects of historical reasoning instruction and writing strategy mastery in culturally and academically diverse middle school classrooms. *Journal of Educational Psychology*, 97(2), 139-156.
- Delpit, L. (1995). *Other people's children: Cultural conflict in the classroom*. New York: The New Press.
- Dewey, J. (1910). *How we think*. New York: D.C. Heath.
- Dewey, J. (1938). *Experience & education*. New York: Simon & Schuster.
- Dimond, S. E. (1948). The Detroit citizenship study. *Social Education*, 12, 356-358.
- Dodge, O. N. (1966). *Generalization and concept development as an instructional method for eighth grade social studies*. (Unpublished doctoral dissertation). Montana State University, Bozeman, MT.
- Doyle, W. (1983). Academic work. *Review of Educational Research*, 53(2), 159-199.
- Education Queensland. (2004). The new basics research report. Retrieved Jan. 15, 2008, from <http://education.qld.gov.au/corporate/newbasics/html/library.html#resreport>
- Elias, G. S. (1958). *An experimental study of teaching methods in ninth grade social studies classes (civics)*. (Unpublished doctoral dissertation). Boston University, Boston, MA.



- Elsmere, R. T. (1961). *An experimental study utilizing the problem-solving approach in teaching United States history* (Unpublished doctoral dissertation). Indiana University, Bloomington, IN.
- Elsmere, R. T. (1963). An experimental study utilizing the problem-solving approach in teaching United States History. *Bulletin of the School of Education Indiana University*, 39(3), 114-139.
- Engle, S. H. (1960). Decision making: The heart of social studies instruction. *Social Education*, 24(6), 301-304, 306.
- Evans, R. W. (2004). *The social studies wars: What should we teach the children?* . New York: Teachers College Press.
- Feilzer, M. Y. (2010). Doing mixed methods research pragmatically: Implications for the rediscovery of pragmatism as a research paradigm. *Journal of Mixed Methods Research*, 4(1), 6-16.
- Fenton, E. (1967). *The new social studies*. New York: Holt, Rinehart and Winston, Inc.
- Ferretti, R. P., MacArthur, C. D., & Okolo, C. M. (2001). Teaching for historical understanding in inclusive classrooms. *Learning Disability Quarterly*, 24, 59-71.
- Finn, C. E., Jr. (2003). Foreward in S.M. Stern, M. Chesson, M.B. Klee, & L. Spoehr (Eds.), *Effective state standards for U.S. history: A 2003 report card* (pp. 5-8): Thomas B. Fordham Institute.

- Foster, S. J., & Yeager, E. A. (1999). "You've got to put together the pieces": English 12-year-olds encounter and learn from historical evidence. *Journal of Curriculum and Supervision, 14*, 286-317.
- Frankville, D. D. (1969). *An evaluation of two methods of teaching American history in grade eleven*. (Unpublished doctoral dissertation). United States International University, San Diego, CA.
- Fraze, B., & Ayers, S. (2003). Garbage in, garbage out: Expanding environments, constructivism, and content knowledge in social studies. In J. Leming, L. Ellington & K. Porter (Eds.), *Where did social studies go wrong?* Washington, D.C.: Thomas B. Fordham Foundation.
- Gabella, M. S. (1994). Beyond the looking glass: Bringing students into the conversation of historical inquiry. *Theory and Research in Social Education, XXII*(3), 340-363.
- Gallagher, S. A., & Stepien, W. J. (1996). Content acquisition in problem-based learning: Depth versus breadth in American studies. *Journal for the Education of the Gifted, 19*(3), 257-275.
- Gaudelli, W. (2006). The future of high-stakes history assessment: Possible scenarios, potential outcomes. In S. G. Grant (Ed.), *Measuring history: Cases of high-stakes testing across the United States*. Greenwich, Conn: Information Age Publishing.
- Glaser, E. M. (1941). *An experiment in the development of critical thinking* (Vol. 843). New York: Teachers College, Columbia University.

- Goodlad, J. (1984). *A place called school: Prospects for the future*. New York: McGraw-Hill.
- Gore, J. M., Ladwig, J. G., Lingard, R., & Luke, A. (2001). Final report of the Queensland school reform longitudinal study. Full report available from Education Queensland.
- Grant, S. G. (2001a). It's just the facts, or is it? The relationship between teacher's practices and students' understandings of history. *Theory and Research in Social Education*, 29(1), 65-108.
- Grant, S. G. (2001b). An uncertain lever: Exploring the influence of state-level testing in New York state on teaching social studies. *Teachers College Record*, 103(3), 398-426.
- Grant, S. G. (2005). More journey than end: A case study of ambitious teaching. In O. L. Davis & E. Yeager (Eds.), *Wise social studies in an age of high-stakes testing* (pp. 117-130). Greenwich, CT: Information Age.
- Grant, S. G., Derme-Insinna, A., Gradwell, J. M., Lauricella, A. M., Pullano, L., & Tzetzko, K. (2002). Juggling two sets of books: A teacher responds to the new global history exam. *Journal of Curriculum and Supervision*, 17(3), 232-255.
- Grant, S. G., & Horn, C. (2006). The state of state-level history testing. In S. G. Grant (Ed.), *Measuring cases: Cases of high-stakes testing across the United States*. Greenwich, Conn.: Information Age Publishing.

- Greene, J. C. (2008). Is mixed methods social inquiry a distinctive methodology? *Journal of Mixed Methods Research*, 2(1), 7-22.
- Greeno, J. G., Collins, A. M., & Resnick, L. B. (1996). Cognition and learning. In D. C. Berliner & R. C. Calfee (Eds.), *Handbook of educational psychology* (pp. 1071). New York: Simon & Schuster Macmillan.
- Gross, R. E., & McDonald, F. (1958). The problem-solving approach. *Phi Delta Kappan*(March), 259-265.
- Hahn, C. L. (1991). Controversial issues in social studies. In J. P. Shaver (Ed.), *Handbook of research on social studies teaching and learning* (pp. 470-479). New York: MacMillan
- Hahn, C. L., & Tocci, C. M. (1990). Classroom climate and controversial issues discussions: A five nation study. *Theory and Research in Social Education*, XVIII(4), 344-362.
- Harmon, L. G. (2006). *The effects of an inquiry-based American history program on the achievement of middle school and high school students*. (Unpublished doctoral dissertation). University of North Texas, Denton, TX.
- Hartzler-Miller, C. (2001). Making sense of "best practice" in teaching history. *Theory and Research in Social Education*, 29(4), 672-695.
- Henderson, K. B. (1958). The teaching of critical thinking. *Phi Delta Kappan*, 39, 280-282.

- Hess, D. (2008a). Controversial issues and democratic discourse. In L. S. Levstik & C. A. Tyson (Eds.), *Handbook of research in social studies education* (pp. 124-136). New York: Routledge.
- Hess, D., & Posselt, J. (2002). How high school students experience and learn from the discussion of controversial issues. *Journal of Curriculum and Supervision*, 17(4), 283-314.
- Hess, F. M. (2008b). *Still at risk: What students don't know, even now*. Washington, DC: Common Core.
- Hirsch, E. D., Jr. (1988). *Cultural literacy: What every American needs to know*. Boston: Houghton Mifflin.
- Hirsch, E. D., Jr. (2009-2010). The anti-curriculum movement: Tragically and unintentionally, it's really an anti-equality movement. *American Educator*, 33(4), 10-11.
- Hunkin, F. P. (1967). *Influence of analysis and evaluation questions on critical thinking and achievement in sixth grade social studies*. (Unpublished doctoral dissertation).
- Hiram, G. A. (1957). Experiment in developing critical thinking in children. *Journal of Experimental Education*, 26.
- Johnson, F. A. (1961). *Depth vs. breadth in teaching American history*. (Unpublished doctoral dissertation). University of Minnesota, Minneapolis, MN.

- Johnson, R. B., Onwuegbuzie, A. J., & Turner, L. A. (2007). Toward a definition of mixed methods research. *Journal of Mixed Methods Research, 1*(2), 112-133.
- Johnston, J., Anderman, E., Milne, L., Klenk, L., & Harris, D. (1994). *Improving civic discourse in the classroom: Taking the measure of Channel One* (Research Report 4). Ann Arbor, MI: University of Michigan.
- Kahne, J., Rodriguez, M., Smith, B. A., & Thiede, K. (2000). Developing citizens for democracy? Assessing opportunities to learn in Chicago's social studies classrooms. *Theory and Research in Social Education, 28*, 318-330.
- Kahne, J. E., & Sporte, S. E. (2008). Developing citizens: The impact of civic learning opportunities on students' commitment to civic participation. *American Educational Research Journal, 45*(3), 738-766.
- Kantrowitz, B., & Wingert, P. (2006, May 8). America's best high schools, 2006. *Newsweek, 147*, 50-54.
- Kight, S. S., & Mickelson, J. M. (1949). Problems vs. subject. *The Clearing House, 24*, 3-7.
- King, M. B., Schroeder, J., & Chawszczewski, D. (2001). Authentic assessment and student performance in inclusive schools, Brief #5, *Research Institute on Secondary Education Reform (RISER) for Youth with Disabilities Brief*. Madison, WI: University of Wisconsin-Madison.

- King, P. M., & Kitchener, K. S. (1994). *Developing reflective judgment: Understanding and promoting intellectual growth and critical thinking in adolescents and adults*. San Francisco, CA: Jossey-Bass Publishers.
- Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist, 41*, 75-86.
- Klentschy, M., Garrison, L., & Amaral, O. (2001). *Valle Imperial Project in Science (VIPS): Four-year comparison of student achievement data, 1995-1999*. El Centro, CA: El Centro School District.
- Koedel, C., & Betts, J. (2009). *Does student sorting invalidate value-added models of teacher effectiveness? An extended analysis of the Rothstein critique*: Department of Economics, University of Missouri.
- Koh, K., Kim, & Luke, A. (2009). Authentic and conventional assessment in Singapore schools: An empirical study of teacher assignments and student work. *Assessment in Education: Principles, Policy, and Practice, 16*(3), 291-318.
- Koh, K., Lee, A. N., Tan, W., Wong, H. M., Guo, L., Lim, T. M., et al. (2005). *Looking collaboratively at the quality of teachers' assessment tasks and student work in Singapore schools*. Paper presented at the International Association for Educational Assessment Conference, Singapore.

- Kohlmeier, J. (2005). The impact of having 9th graders "do history". *The History Teacher*, 38(4), 499-524.
- Kohlmeier, J. (2006). "Couldn't she just leave?": The relationship between consistently using class discussions and the development of historical empathy in a 9th grade world history course. *Theory and Research in Social Education*, 34(1), 34-57.
- Kornhaber, M. L. (2004). Appropriate and inappropriate forms of testing, assessment, and accountability. *Educational Policy*, 18(1), 45-70.
- Kozma, R. B. (2008). 21st Century Skills, education, and competitiveness: A resource and policy guide. Retrieved May 15, 2009, from [http://www.txccrs.org/downloads/Partnership\\_21stCenturySkills.pdf](http://www.txccrs.org/downloads/Partnership_21stCenturySkills.pdf)
- Ladwig, J. G., Smith, M., Gore, J., Amosa, W., & Griffiths, T. (2007). *Quality of pedagogy and student achievement: Multi-level replication of authentic pedagogy*. Paper presented at the Australian Association for Research in Education Conference, Fremantle.
- Lake Corporate Consulting. (2006). Standardised literacy and numeracy scores and 'doing' the New Basics. Retrieved Feb. 18, 2008, from [http://education.qld.gov.au/corporate/newbasics/pdfs/litnum\\_rpt.pdf](http://education.qld.gov.au/corporate/newbasics/pdfs/litnum_rpt.pdf)
- Lambert, R. A. (1980). Effects of moral education strategies on increased subject matter content of secondary school social studies students. (Unpublished doctoral dissertation). Catholic University of America, Washington, D.C..



- Larson, B. E. (2003). Comparing face-to-face discussion and electronic discussion: A case study from high school social studies. *Theory and Research in Social Education, 31*(3), 347-365.
- Laws of Florida. (2006). Ch. 2006-74 (House Bill 7087), item 1003.42.2.f, signed June 5, 2006. from [http://laws.flrules.org/files/Ch\\_2006-074.pdf](http://laws.flrules.org/files/Ch_2006-074.pdf)
- Lee, M. A. (1967). *Development of inquiry skills in ungraded social studies classes in a junior high school* (Unpublished doctoral dissertation). Indiana University, Bloomington, IN.
- Lee, P., & Ashby, R. (2000). Progression in historical understanding among students ages 7-14. In P. N. Stearns, P. Seixas & S. Wineburg (Eds.), *Knowing, teaching, and learning history: National and international perspectives* (pp. 199-222). New York: New York University Press.
- Lee, V. E., & Smith, J. B. (1995). Effects of high school restructuring and size on early gains in achievement and engagement. *Sociology of Education, 68*(4), 241-270.
- Lee, V. E., Smith, J. B., & Croninger, R. G. (1997). How high school organization influences the equitable distribution of learning in science and mathematics. *Sociology of Education, 70*(April), 128-150.
- Lee, V. E., Smith, J. B., & Newmann, F. M. (2001). *Instruction and achievement in Chicago elementary schools*. Chicago, IL: Consortium on Chicago School Research.

- Leming, J. S. (2003). Ignorant Activists: Social change, "higher order thinking," and the failure of social studies. In J. Leming, L. Ellington & K. Porter (Eds.), *Where Did Social Studies Go Wrong?* (pp. 124-142). Washington, D.C.: Thomas B. Fordham Foundation.
- Levin, M., Newmann, F. M., & Oliver, D. (1969). *A law and social science curriculum based on the analysis of public issues* (No. Final Report project no. HS 058. Grant no. OE 310142). Washington, D.C.: Department of Health, Education, and Welfare.
- Levstik, L. S. (2008). What happens in social studies classrooms? Research on K-12 social studies practice. In L. S. Levstik & C. A. Tyson (Eds.), *Handbook of research in social studies education*. New York, NY: Taylor and Francis.
- Lipka, R. P., Lounsbury, J. H., Toepfer, C. F., Jr., Vars, G. F., Alessi, S. P., & Kridel, C. (1998). *The Eight-Year Study revisited: Lessons from the past for the present*. Columbus, OH: National Middle School Association.
- Mackenzie, A. W., & White, R. T. (1982). Fieldwork in geography and long-term memory structures. *American Educational Research Journal*, 19, 623-632.
- Madden, J. R. (1970). *The relationship between the use of an inquiry teaching technique in a social studies classroom and the attitude of students toward the social studies course*. (Unpublished doctoral dissertation). Syracuse University, Syracuse, NY.

- Massialas, B. G. (1961). *Description and analysis of teaching a high school course in World History* (Unpublished doctoral dissertation). Indiana University, Bloomington, IN.
- Massialas, B. G. (1963). The Indiana experiments in inquiry: Social studies. *Bulletin of the School of Education, Indiana University*, 39(3).
- Massialas, B. G., & Cox, C. B. (1966). *Inquiry in social studies*. New York: McGraw-Hill Book Company.
- McDevitt, M., & Kiouisis, S. (2006). Experiments in political socialization: Kids Voting USA as a model for civic education reform. *Circle Working Paper 49*.
- McNeil, L. (1986). *Contradictions of control: School structure and school knowledge*. New York: Routledge and Kegan Paul.
- McNeil, L., & Valenzuela, A. (2001). The harmful impact of the TAAS system of testing in Texas: Beneath the accountability rhetoric. In G. Orfield & M. L. Kornhaber (Eds.), *Raising Standards or Raising Barriers? Inequality and High-Stakes Testing in Public Education* (pp. 127-150). New York: Century Foundation Press.
- Metcalf, L. E. (1963). Research on teaching the social studies. In N. L. Gage (Ed.), *Handbook of Research on Teaching*. Chicago, IL: Rand McNally & Company.
- Monte-Sano, C. (2008). Qualities of historical writing instruction: A comparative case study of two teachers' practices. *American Educational Research Journal*, 45(4), 1045-1079.

Morton, J. B. (2004). *Alabama course of study: Social studies* (Bulletin 2004, No. 18): Alabama Department of Education.

Morton, J. B. (2009). The handbook of administrative procedures for the Alabama High School Graduation Exam. Retrieved May 17, 2010, from <https://docs.alsde.edu/documents/91/Handbook%20of%20Administrative%20Procedures%20for%20the%20AHSGE%202009.pdf>

Nash, G. (1995). The history children should study. *Chronicle of Higher Education*, *XLI*(32), A60.

National Council for the Social Studies. (1994). *Expectations of excellence: Curriculum standards for social studies*. Silver Spring, MD: National Council for the Social Studies.

Newmann, F. M. (1990). A test of higher order thinking in social studies: Persuasive writing on constitutional issues using the NAEP approach. *Social Education*, *54*, 369-373.

Newmann, F. M. (1991a). Classroom thoughtfulness and students' higher order thinking: Common indicators and diverse social studies courses. *Theory and Research in Social Education*, *XIX*(4), 410-433.

Newmann, F. M. (1991b). Higher order thinking in the teaching of social studies: Connections between theory and practice. In J. F. Voss, D. N. Perkins & J. W. Segal (Eds.), *Informal reasoning and education*. Mahwah, NJ: Lawrence Erlbaum.

- Newmann, F. M. (1991c). Promoting higher order thinking in social studies: Overview of a study of 16 high school departments. *Theory and Research in Social Education, XIX*(4), 324-340.
- Newmann, F. M., & Archbald, D. A. (1988). The functions of assessment and the nature of authentic academic achievement. In A. Berlak (Ed.), *Assessing achievement: Toward the development of a new science of educational testing*. Buffalo, NY: SUNY.
- Newmann, F. M., & Associates. (1996). *Authentic achievement: Restructuring schools for intellectual quality*. San Francisco: Jossey-Bass.
- Newmann, F. M., Bryk, A. S., & Nagaoka, J. K. (2001). *Authentic intellectual work and standardized tests: Conflict or coexistence?* Chicago: Consortium on Chicago School Research.
- Newmann, F. M., King, M. B., & Carmichael, D. L. (2007). *Authentic instruction and assessment: Common strategies for rigor and relevance in teaching academic subjects*: Prepared for the Iowa Department of Education.
- Newmann, F. M., Lopez, G., & Bryk, A. S. (1998). *The quality of intellectual work in Chicago Schools: A baseline report*. Chicago: Consortium on Chicago School Research.
- Newmann, F. M., Marks, H. M., & Gamoran, A. (1996). Authentic pedagogy and student performance. *American Journal of Education, 104*(4), 280-312.

- Newmann, F. M., & Oliver, D. (1970). *Clarifying public controversy: An approach to social studies*. Boston: Little, Brown.
- Newmann, F. M., Secada, W. G., & Wehlage, G. G. (1995). *A guide to authentic instruction and assessment: Vision, standards, and scoring*. Madison: Center on Organization and Restructuring of Schools, Wisconsin Center for Education Research, University of Wisconsin.
- Newmann, F. M., Wehlage, G. G., & Lamborn, S. D. (1992). The significance and sources of student engagement. In F. Newmann (Ed.), *Student Engagement and Achievement in American Secondary Schools* (pp. 11-39). New York: Teachers College Press.
- No Child Left Behind (NCLB) Act of 2001, 20 U.S.C.A. & 6301 *et seq.* (West 2003)
- Noel, R. C. (1996). The "authentic pedagogy" study. Review No. One. Retrieved May 14, 2006, from <http://www.mathematically.correct.com/qed.htm>
- Nuthall, G., & Alton-Lee, A. (1995). Assessing classroom learning: How students use their knowledge and experience to answer classroom achievement test questions in science and social studies. *American Educational Research Journal*, 32(1), 185-223.
- Nystrand, M., & Gamoran, A. (1990). *Student engagement: When recitation becomes conversation* (No. ED 323 581). Madison, WI: National Center on Effective Secondary Schools.

- Oakes, J. (2005). *Keeping track: How schools structure inequality* (2nd ed.). New Haven, CT: Yale University Press.
- Oliver, D., & Shaver, J. P. (1966). *Teaching public issues in the high school*. Boston: Houghton Mifflin.
- Onosko, J. J. (1991). Barriers to the promotion of higher order thinking in social studies. *Theory and Research in Social Education*, XIX(4), 341-366.
- Osborne, J., & Waters, E. (2002). Four assumptions of multiple regression that researchers should always test [Electronic Version]. *Practical Assessment, Research, and Evaluation*, 8. Retrieved Sept. 7, 2008 from <http://pareonline.net/getvn.asp?v=8&n=2>
- Parker, W. C. (1991). Achieving thinking and decision-making objectives in social studies. In J. P. Shaver (Ed.), *Handbook of research on social studies teaching and learning*. New York: Macmillan.
- Parker, W. C. (1996). Introduction: Schools as laboratories of democracy. In W. Parker (Ed.), *Educating the Democratic Mind* (pp. 1-22). New York: State University of New York Press.
- Parker, W. C., Mueller, M., & Wendling, L. (1989). Critical reasoning on civic issues. *Theory and Research in Social Education*, 17(1), 7-32.
- Partnership for 21st Century Skills. (2007). *Beyond the three Rs: Voter attitudes toward 21st Century Skills*. Tucson, AZ: Partnership for 21st Century Skills.

- Patton, M. Q. (1987). *How to use qualitative methods in evaluation*. Newbury Park, CA: Sage Publications.
- Paul Gagnon and the Bradley Commission on History in Schools (Ed.). (1989). *Historical literacy: The case for history in American education*. New York: Macmillan.
- Peters, C. C. (1948). *Teaching high school history and social studies for citizenship training: The Miami experiment in democratic, action-centered education*. Coral Gables, FL: University of Miami bookstore.
- Piaget, J. (1952). *The origins of intelligence in children* (M. Cook, Trans.). New York: International Universities Press, Inc.
- Pink, D. (2008). Tom Friedman on education in the 'flat world': A discussion with author Daniel Pink on curiosity, passion and the politics of school reform in the global marketplace (Interview). *School Administrator*, 65(2), 12.
- Quillen, I. J., & Hanna, L. A. (1948). *Education for social competence*. Chicago, IL: Scott Foresman.
- Ravitch, D., & Finn, C. E. (1987). *What do our 17-year-olds know?* New York: Harper & Row Publishers.
- Rehage, K. J. (1951). A comparison of pupil-teacher and teacher-directed procedures in eighth grade social studies classes. *Journal of Educational Research*, 45, 111-115.
- Resnick, L. B. (1987). Learning in school and out. *Educational Researcher*, 16(9), 13-20.



- Richardson, E. (2000). *Social studies items specifications for the Alabama High School Graduation Exam* (Bulletin 2000, No. 49): Alabama Department of Education.
- RISER. (2000). *Authentic instruction scoring manual*. Madison, WI: Wisconsin Center for Education Research.
- Roelofs, E., & Terwel, J. (1999). Constructivism and authentic pedagogy: State of the art and recent developments in the Dutch national curriculum in secondary education. *Journal of Curriculum Studies*, 31(2), 201-227.
- Rogers, C., & Freiburg, H. J. (1994). *Freedom to learn*. New York: Macmillan College Publishing Company.
- Rose, H. P. (1970). *The relationship between methods used to teach American history and changes in attitude and achievement*. (Unpublished doctoral dissertation). United International University.
- Rossi, J. A. (1995). In-depth study in an issues-oriented social studies classroom. *Theory and Research in Social Education*, XXIII(2), 88-120.
- Rossi, J. A. (1998). Issues-centered instruction with low-achieving high school students: The dilemmas of two teachers. *Theory and Research in Social Education*, 26(3), 380-409.
- Rothstein, A. (1960). *An experiment in developing critical thinking through the teaching of American history*. (Unpublished doctoral dissertation). New York University, New York.

- Rothstein, R. (2004). We are not ready to assess history performance [Electronic Version]. *The Journal of American History*, 90. Retrieved Dec. 16, 2008 from <http://www.historycooperative.org>.
- Rothstein, R. (2009). Replacing No Child Left Behind [Electronic Version]. *Education Week*, 28, 28-29. Retrieved August 11, 2009.
- Rumelhart, D. E. (1980). Schemata: The building blocks of cognition. In R. J. Spiro, B. C. Bruce & W. F. Brewer (Eds.), *Theoretical issues in reading comprehension*. Hillsdale, NJ: Lawrence Erlbaum.
- Saxe, D. W. (2003). Patriotism versus multiculturalism in times of war. *Social Education*, 67(2), 107-109.
- Saye, J. W., & Brush, T. (1999a). Student engagement with social issues in a multimedia-supported learning environment. *Theory and Research in Social Education*, 27(4), 472-504.
- Saye, J. W., & Brush, T. (1999b). *Student reasoning about ill-structured social problems in a multimedia-supported learning environment*. Paper presented at the Annual Meeting of the National Council for the Social Studies, Orlando, FL.
- Saye, J. W., & Brush, T. (2002). Scaffolding critical reasoning about history and social issues in multimedia-supported learning environments [Electronic Version]. *Educational Technology Research and Development*, 50, 77-96.

- Saye, J. W., & Brush, T. (2004). Promoting civic competence through problem-based history learning environments. In G. E. Hamot, J. J. Patrick & R. S. Leming (Eds.), *Civic learning in teacher education: International perspectives on education for democracy in the preparation of teachers* (Vol. 3, pp. 123-145). Bloomington, Indiana: ERIC Clearinghouse for Social Studies/ Social Science Education.
- Saye, J. W., & Brush, T. (2007). Using technology-enhanced learning environments to support problem-based historical inquiry in secondary school classrooms. *Theory and Research in Social Education*, 35(2), 196-230.
- Scheurman, G. (1998). From behaviorist to constructivist teaching. *Social Education*, 62, 6-9.
- Schroeder, J. L., Braden, J. P., & King, B. (2001). *Standards and scoring criteria for assessment tasks and student performance*. Madison: Research Institute on Secondary Education Reform for Youth with Disabilities.
- Schug, M. C. (2003). Teacher-centered instruction: The Rodney Dangerfield of social studies. In J. S. Leming, L. Ellington & K. Porter (Eds.), *Where did social studies go wrong?* Washington, D.C.: Thomas B. Fordham Foundation.
- Seixas, P. (2001). Review of research on social studies. In V. Richardson (Ed.), *Handbook of research on teaching*. Washington, D.C.: American Educational Research Association.
- Sizer, T. R. (1984). *Horace's compromise*. Boston: Houghton Mifflin.

- Smith, J., & Niemi, R. G. (2001). Learning history in school: The impact of course work and instructional practices on achievement. *Theory and Research in Social Education*, 29(1), 18-42.
- Social Studies Inquiry Research Collaborative. (2011). *Authentic pedagogy: Examining intellectual challenge in a national sample of social studies classrooms*. Paper presented at the Annual Meeting of the American Education Research Association Conference, New Orleans, LA.
- Stecher, B. M. (2002). Consequences of large-scale, high-stakes testing on school and classroom practice. In L. S. Hamilton, B. M. Stecher & S. P. Klein (Eds.), *Making sense of test-based accountability in education*. Santa Monica, CA: RAND.
- Stern, S. M., Chesson, M., Klee, M. B., & Spoehr, L. (2003). *Effective state standards for U.S. History: A 2003 report card*: Thomas B. Fordham Institute.
- Stevens, J. (2002). *Applied multivariate statistics for the social sciences* (4th ed.). Mahwah, NJ: Erlbaum.
- Stewart, B. E. (2006). *Value added modeling: The challenge of measuring educational outcomes*. New York: Carnegie Corporation of New York.
- Stewart, R. A., & Brendefur, J. L. (2005). Fusing lesson study and authentic achievement: A model for teacher collaboration. *Phi Delta Kappan*, 681-687.
- Stiggins, R. J., & Conklin, N. F. (1992). *In teachers' hands: Investigating the practices of classroom assessment*. New York: State University of New York Press.

- Symcox, L. (2002). *Whose history? The struggle for national standards in American classrooms*. New York, NY: Teacher's College Press.
- Taba, H. (1966). *Teaching strategies and cognitive functioning in elementary school children* (Cooperative Research Project No. 2404). San Francisco: San Francisco State College.
- Taba, H., Levine, S., & Elzey, F. F. (1964). *Thinking in elementary school children* (Cooperative Research Project No. 1574). San Francisco, CA: San Francisco State College.
- Terwilliger, J. S. (1997). Semantics, psychometrics, and assessment reform: A close look at "authentic" assessments. *Educational Researcher*, 26(8), 24-27.
- Terwilliger, J. S. (1998). Rejoinder: Response to Wiggins and Newmann. *Educational Researcher* (August-September), 22-23.
- Thompson, S. (2001). The authentic standards movement and its evil twin. *Phi Delta Kappan*, 82(5).
- Thornton, S. J. (1991). Teacher as curricular-instructional gatekeeper in social studies. In J. P. Shaver (Ed.), *Handbook of research on social studies teaching and learning* (pp. 237-248). New York: Macmillan.
- Torney-Purta, J. (2002). The school's role in developing civic engagement: A study of adolescents in twenty-eight countries. *Applied Developmental Science*, 6(4), 203-212.

- Trochim, W. M. (2006). The research methods knowledge base. 2nd edition. from <http://www.socialresearchmethods.net/kb/> (version current as of October 20, 2006).
- U.S. Census Bureau. (2000a). Profile of selected economic characteristics. Retrieved Sept. 15, 2009, from <http://censtats.census.gov/data/AL/1600103076.pdf>
- U.S. Census Bureau. (2000b). State and county quickfacts. Retrieved September 15, 2009, from <http://quickfacts.census.gov/qfd/states/01/0103076.html>
- VanSickle, R. L., & Hoge, J. D. (1991). Higher cognitive thinking skills in social studies: Concepts and critiques. *Theory and Research in Social Education, 19*(2), 152-172.
- VanSledright, B. A. (2002). Fifth graders investigating history in the classroom: Results from a researcher-practitioner design experiment. *Elementary School Journal, 102*, 131-160.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.
- Wallen, N. E., & Travers, R. M. W. (1963). Analysis and investigation of teaching methods. In N. L. Gage (Ed.), *Handbook of research on teaching*. Chicago, IL: Rand McNally & Company.
- Wallis, C., & Steptoe, S. (2006, December 18). How to bring our schools out of the 20th century. *Time, 50-56*.

- Wenzel, S., Nagaoka, J. K., Morris, L., Billings, S., & Fendt, C. (2002). *Documentation of the 1996-2002 Chicago Annenberg research strand on authentic intellectual demand exhibited in assignments and student work: A technical process manual*. Chicago: Consortium of Chicago School Research.
- Whitehead, A. N. (1929). *The aims of education and other essays*. New York: Macmillan.
- Wiggins, G. (1989). A true test: Toward more authentic and equitable assessment. *Phi Delta Kappan*, 70(9), 703-713.
- Wiggins, G. (1993a). *Assessing student performance*. San Francisco, CA: Jossey-Bass.
- Wiggins, G. (1993b). Assessment to improve performance, not just monitor it: Assessment reform in the social sciences. *Social Science Record*(Fall), 5-12.
- Williams, J. M. (1981). *A comparison study of the effects of inquiry and traditional teaching procedures on student attitude, achievement, and critical-thinking ability in eleventh grade United States history*. (Unpublished doctoral dissertation). Auburn University, Auburn, AL.
- Williamson, J. L. (1966). *The effectiveness of two approaches to the teaching of high school American history*. (Unpublished doctoral dissertation). University of North Texas, Denton, TX.

- Windschitl, M. (2002). Framing constructivism in practice as the negotiation of dilemmas: An analysis of the conceptual, pedagogical, cultural, and political challenges facing teachers. *Review of Educational Research*, 72(2), 131-175.
- Wineburg, S. (1991). On the reading of historical texts: Notes on the breach between school and academy. *American Educational Research Journal*, 28, 495-519.
- Wineburg, S. (2001). *Historical thinking and other unnatural acts: Charting the future of teaching the past*. Philadelphia: Temple University Press.
- Womack, J. A. (1969). *An analysis of inquiry-oriented high school geography project urban materials*. (Unpublished doctoral dissertation), United States International University, San Diego, CA.
- Yost, D. E. (1972). *The effect of two instructional methods on achievement, critical thinking, and study habits and attitudes in tenth grade American government classes*. (Unpublished doctoral dissertation). University of Northern Colorado, Greeley, CO.
- Young, K. M., & Leinhardt, G. (1998). Writing from primary documents: A way of knowing history. *Written Communication*, 15, 25-68.



## **APPENDICES**

Appendix A: Teacher Interview Script

**Teacher Background Data:**

1. What is your gender?

- Male
- Female

2. What is your age?

- 25 or less
- between 26 and 35
- between 36 and 45
- between 46 and 55
- greater than 55

3. What is your ethnicity?

- African American
- Asian American
- Latino/Hispanic American
- Native American
- White (other than Latino)
- Other/No Response

4. How many years have you been a teacher?

- 2 years or less
- between 3 and 5 years
- between 6 and 10 years
- between 11 and 15 years
- greater than 15 years

(5) How many years have you been teaching at this particular school?

(6) How many years have you been teaching a course that is offered at the grade level in which the state-mandated exam is administered?

(7) Are you National Board certified?      Highly Qualified?

(8) What is your schedule of classes?

## Teacher Background data (continued)

- (9) What is the highest degree you attained?
- (10) What was your major in college? Did you have a concentration in one or more of the social science disciplines? Did you have some type of alternative certification (i.e. 5<sup>th</sup> year program, Teach for America, etc.)?

## Challenging Tasks

(Partnership for 21st Century Skills) Why do you view these tasks as the most challenging for students?

(Partnership for 21st Century Skills) Can you provide a description of what you want students to do for these tasks?

### Task 1:

*Goal(s) for student learning:*

*Where does the assignment fall in the semester (context)?*

*Is this assignment modified in any way for other blocks? Do all blocks do the assignment?*

Same questions for Tasks 2 and 3.

## Graduation Exam

(Partnership for 21st Century Skills) Do you incorporate activities that are explicitly focused on preparation for a high-stakes exam?

(Partnership for 21st Century Skills) What types of materials do you use to prepare students for the graduation exam (i.e. practice test booklets, computer drill & practice programs, transparencies or powerpoints with specific questions, etc.)?

(3) Provide an estimate of the total amount of time you spend each semester on graduation exam preparation.

- No more than 1 day  
 2 to 4 days  
 1 week  
 Over 1 to 3 weeks  
 1 to 2 months  
 Over 2 to 3 months

\_\_\_ Over 3 to 6 months  
\_\_\_ Over 6 months

(4) Does your graduation exam preparation vary from class to class? Which classes receive more explicit preparation? Why?

## Appendix B: Teacher Recruitment Script

Hello, my name is Lamont Maddox, and I am a doctoral student from Auburn University. I would like to invite you to participate in a study that is focused on the following question: **How does the kind of learning that students experience in their social studies classes affect their performances on both low and high order assessments?** I am trying to gain a better understanding of what works to improve performance for different groups of students (i.e. low SES, male/female, ethnicity, etc.). You are being asked to participate because you teach a social studies course on either the 9<sup>th</sup> or 10<sup>th</sup> grade level.

As part of this study, I would like to analyze the nature of the instructional experiences tenth grade students encounter during the social studies classes they take just prior to their first attempt on the Alabama High School Graduation Exam (AHSGE). Each teacher takes a unique approach to his/her instruction. I want to use three classroom observations and an analysis of three challenging assignments in each class as a means to better understand the social studies experiences that Auburn students have. Additional information regarding the context of these assignments will be gained through a brief interview. I would like to emphasize that I am not developing a rating that judges one form of instruction as better than another. I am cataloging the type of work students are asked to do in their classes. A one size fits all approach may very well not work with all students. I want to see which instructional experiences produce results with which groups of students.

Student achievement data for this study is coming from two sources. One source is the AHSGE. The other source is an essay being piloted by the school system. This essay is being used as the higher order assessment of what students are able to retain and do following a previous unit of instruction. Participating teachers will teach and assess the designated focus unit as they always do. The essay will be administered at a later date and then the school system will provide the essays to me for independent scoring. Student names will be removed from all student data and replaced with code numbers prior to its delivery to me.

This study will be conducted during the Spring 08, Fall 08, and Spring 09 semesters. However, the test of higher order thinking will only be administered to one group of tenth graders during the Spring 08 semester. Therefore, all teacher participation should be complete prior to the end of this school year. I estimate that the study will only require a couple of hours of your time beyond the scope of your normal day to day responsibilities. If you agree to participate, the study will have a minimal impact on the regular instruction you provide your classes. Throughout this process, teacher confidentiality will be protected. I will not link teacher names to data, but will describe how various classroom experiences are correlated to student outcomes. I will also track student experiences over multiple years (9<sup>th</sup> & 10<sup>th</sup> grade) rather than with a single teacher. This will provide further support in maintaining teacher confidentiality.

The results of this study could help you to make decisions about how to reach the needs of all of your students. I hope you will join me in this study. Any questions?

## Appendix C: Scoring Criteria for Classroom Instruction

Scoring instructions: To determine scores for the four standards, follow the technical scoring criteria as outlined in the tips below. Consider the descriptions for scores 1-5 on each standard to constitute the *minimum* criteria for that score. If you find yourself between scores, make the decision by asking whether the minimum conditions of the higher score have been met. If not, use the lower score. In determining scores for each standard, the observer should consider only the evidence observed during the lesson observation. “Many” students refers to at least 1/3 of the students in a class; “most” refers to more than half; “almost all” is not specified numerically, but should be interpreted as “all but a few.”

Date: \_\_\_\_\_

Class Observed: \_\_\_\_\_

Observer: \_\_\_\_\_

Score	HOTS					Deep Knowledge								
	<i>To what extent do students use lower order thinking processes? To what extent do students use higher order thinking processes?</i>					<i>To what extent is knowledge deep? To what extent is knowledge shallow and superficial?</i>								
	<i>Lower Order Thinking Only</i>	1	2	3	4	5	<i>Higher Order Thinking Only</i>	<b>Knowledge is shallow</b>	1	2	3	4	5	<b>Knowledge is deep</b>
<b>5</b>	Almost all students, almost all of the time, are performing HOT.					Knowledge is very deep because the teacher successfully structures the lesson so that almost all students sustain a focus on a significant topic and do at least one of the following: demonstrate their understanding of the problematic nature of information and/or ideas; demonstrate complex understanding by arriving at a reasoned, supported conclusion; or explain how they solved a complex problem. In general, students' reasoning, explanations and arguments demonstrate fullness and complexity of understanding.								
<b>4</b>	Students are engaged in at least one major activity during the lesson in which they perform HOT operations, and this activity occupies a substantial portion ( <b>at least 1/3</b> ) of the lesson and many students are performing HOT.					Knowledge is relatively deep because either the teacher or the students provide information, arguments or reasoning that demonstrate the complexity of an important idea. The teacher structures the lesson so that <b>many</b> students <b>sustain</b> a focus on a significant topic for a period of time <b>and do</b> at least one of the following: demonstrate their understanding of the problematic nature of information and/or ideas; demonstrate understanding by arriving at a reasoned, supported conclusion; or explain how they solved a relatively complex problem.								
<b>3</b>	Students are primarily engaged in routine LOT operations a good share of the lesson. There is at least one significant question or activity in which some students perform some HOT operations.					Knowledge is treated unevenly during instruction; i.e., deep understanding of something is countered by superficial understanding of other ideas. At least one significant idea may be presented in depth and its significance grasped, but in general the focus is not sustained.								
<b>2</b>	Students are primarily engaged in LOT, but at some point they perform HOT as a minor diversion within the lesson.					Knowledge remains superficial and fragmented; while some key concepts and ideas are mentioned or covered, only a superficial acquaintance or trivialized understanding of these complex ideas is evident.								
<b>1</b>	Students are engaged only LOT operation; i.e., they either receive, or recite, or participate in routine practice and in no activities during the lesson do students go beyond LOT.					Knowledge is very thin because it does not deal with significant topics or ideas; teacher and students are involved in the coverage of simple information which they are to remember.								

## Scoring Criteria for Classroom Instruction (Cont.)

Score	<p style="text-align: center;"><b>Substantive Conversation</b></p> <p><i>To what extent is classroom discourse devoted to creating or negotiating understandings of subject matter?</i></p> <p style="text-align: center;"> <b>no substantive conversation</b>      <b>high level substantive conversation</b>                      1   2   3   4   5                 </p>	<p style="text-align: center;"><b>Connectedness to the Real World</b></p> <p><i>To what extent is the lesson, activity, or task connected to competencies or concerns beyond the classroom?</i></p> <p style="text-align: center;"> <b>no connection</b>      1   2   3   4   5      <b>connected</b> </p>
<b>5</b>	All features of substantive conversation occur, with at least one example of <u>sustained</u> conversation, and almost all students participate.	Students study or work on a topic, problem or issue that the teacher and students see as connected to their personal experiences or actual contemporary or persistent public issues. Students recognize the connection between classroom knowledge and situations outside the classroom. They explore these connections in ways that create personal meaning and significance for the knowledge. This meaning and significance is strong enough to lead students to become involved in an effort to affect or influence a larger audience beyond their classroom in one of the following ways: by communicating knowledge to others (including within the school), advocating solutions to social problems, providing assistance to people, creating performances or products with utilitarian or aesthetic value.
<b>4</b>	All features of substantive conversation occur, with at least one example of <u>sustained</u> conversation, and many students participate in some substantive conversation (even if not part of the sustained conversation).	Students study or work on a topic, problem or issue that the teacher and students see as connected to their personal experiences or actual contemporary or persistent public issues. Students recognize the connection between classroom knowledge and situations outside the classroom. They explore these connections in ways that create personal meaning and significance for the knowledge. However, there is no effort to use the knowledge in ways that go beyond the classroom to actually influence a larger audience.
<b>3</b>	Substantive Conversation Feature # 2 (sharing) and/or #3 (coherent promotion of collective understanding) occur and involve at least one example of <u>sustained</u> conversation (i.e., at least 3 consecutive interchanges).	Students study a topic, problem or issue that the teacher succeeds in connecting to students' actual experiences or to actual contemporary or persistent public issues. Students recognize some connection between classroom knowledge and situations outside the classroom, but they do not explore the implications of these connections which remain abstract or hypothetical. There is no effort to actually influence a larger audience.
<b>2</b>	Substantive Conversation Feature # 2 (sharing) and/or # 3 (coherent promotion of collective understanding) occur briefly and involve at least one example of two consecutive interchanges.	Students encounter a topic, problem or issue that the teacher tries to connect to students' experiences or to actual contemporary or persistent public issues; i.e., the teacher informs students that there is potential value in the knowledge being studied because it relates to the world beyond the classroom. For example, students are told that understanding Middle East history is important for politicians trying to bring peace to the region; however, the connection is weak and there is no evidence that students make the connection.
<b>1</b>	Virtually no features of substantive conversation occur during the lesson.	Lesson topic and activities have no clear connection to anything beyond itself; the teacher offers no justification beyond the need to perform well in class.



## Appendix D: Scoring Tips for Instruction Rubric

<b>Tips for Scoring HOTS</b>	<b>Tips for Scoring Deep Knowledge</b>
<ul style="list-style-type: none"><li>• Lower order thinking (LOT) occurs when students are asked to receive or recite factual information or to employ rules and algorithms through repetitive routines. As information receivers, students are given pre-specified knowledge ranging from simple facts and information to more complex concepts. Such knowledge is conveyed to students through a reading, work sheet, lecture or other direct instructional medium. Students are not required to do much intellectual work since the purpose of the instructional process is to simply transmit knowledge or to practice procedural routines. Students are in a similar role when they are reciting previously acquired knowledge; i.e., responding to test-type questions that require recall of pre-specified knowledge. More complex activities still may involve LOT when students only need to follow pre-specified steps and routines or employ algorithms in a rote fashion.</li><li>• Higher order thinking (HOT) requires students to manipulate information and ideas in ways that transfer their meaning and implications. This transformation occurs when students combine facts and ideas in order to synthesize, generalize, explain, hypothesize or arrive at some conclusion or interpretation. Manipulating information and ideas through these processes allows students to solve problems and discover new (for them) meanings and understandings.</li><li>• When students engage in HOT, an element of uncertainty is introduced into the instructional process and makes instructional outcomes not always predictable; i.e., the teacher is not certain what will be produced by students. In helping students become producers of knowledge, the teacher's main instructional task is to create activities or environments that allow them opportunities to engage in HOT.</li></ul>	<ul style="list-style-type: none"><li>• Knowledge is shallow, thin or superficial when it does not deal with significant concepts or central ideas of a topic or discipline. Knowledge is also shallow when important, central ideas have been trivialized, or when it is presented as non-problematic. Knowledge is thin when students' understanding of important concepts or issues is superficial such as when ideas are covered in a way that gives them only a surface acquaintance with their meaning. This superficiality can be due, in part, to instructional strategies such as when teachers cover large quantities of fragmented ideas and bits of information that are unconnected to other knowledge.</li><li>• Evidence of shallow understanding by students exists when they do not or can not use knowledge to make clear distinctions, arguments, solve problems and develop more complex understanding of other related phenomena.</li><li>• Knowledge is deep or thick when it concerns the central ideas of a topic or discipline and because such knowledge is judged to be crucial to a topic or discipline.</li><li>• For students, knowledge is deep when they develop relatively complex understandings of these central concepts. Instead of being able to recite only fragmented pieces of information, students develop relatively systematic, integrated or holistic understanding. Mastery is demonstrated by their success in producing new knowledge by discovering relationships, solving problems, constructing explanations, and drawing conclusions.</li><li>• In scoring this item, observers should note that depth of knowledge and understanding refers to the substantive character of the ideas that the teacher presents in the lesson, or to the level of understanding that students demonstrate as they consider these ideas. It is possible to have a lesson that contains substantively important, deep knowledge, but students do not become engaged or they fail to show understanding of the complexity or the significance of the ideas. Observers' ratings can reflect either the depth of the teacher's knowledge or the depth of understanding that students develop of that content.</li></ul>

### Tips for Scoring Substantive Conversation

- This scale measures the extent of talking to learn and to understand in the classroom. There are two dimensions to this construct: one is the substance of subject matter, and the other is the character of dialogue.
- In classes where there is little or no substantive conversation, teacher-student interaction typically consists of a lecture with recitation where the teacher deviates very little from delivering a preplanned body of information and set of questions; students typically give very short answers. Because the teacher's questions are motivated principally by a preplanned checklist of questions, facts, and concepts, the discourse is frequently choppy, rather than coherent; there is often little or no follow-up of student responses. Such discourse is the oral equivalent of fill-in-the-blank or short-answer study questions.
- In classes characterized by high levels of substantive conversation there is considerable teacher-student and student-student interaction about the ideas of a topic; the interaction is reciprocal, and it promotes coherent shared understanding. (Partnership for 21st Century Skills) The talk is about subject matter in the discipline and includes higher order thinking such as making distinctions, applying ideas, forming generalizations, raising questions; not just reporting of experiences, facts, definitions, or procedures. (Partnership for 21st Century Skills) The conversation involves sharing of ideas and is not completely scripted or controlled by one party (as in teacher-led recitation). Sharing is best illustrated when participants explain themselves or ask questions in complete sentences, and when they respond directly to comments of previous speakers. (3) The dialogue build coherently on participants' ideas to promote improved collective understanding of a theme or topic (which does not necessarily require an explicit summary statement). In short, substantive conversation resembles the kind of sustained exploration of content characteristic of a good seminar where student contributions lead to shared understandings.
- To recognize sustained conversations, we define an interchange as a statement by one person and a response by another. Interchanges can occur between teacher and student or student and student. Sustained conversation is defined as at least three consecutive interchanges. The interchanges need not be between the same two people, but they must be linked substantively as consecutive responses. Consecutive responses should demonstrate sensitivity either by responding directly to the ideas of another speaker or by making an explicit transition that shows the speaker is aware he/she is shifting the conversation. Substantive conversation includes the 3 features described above. Each of the features requires interchange between two or more people. None can be illustrated through monologue by one person.

### Tips for Scoring Value Beyond School

- This scale measures the extent to which the class has value and meaning beyond the instructional context. In a class with little or no value beyond, activities are deemed important for success only in school (now or later), but for no other aspects of life. Student work has no impact on others and serves only to certify their level of competence or compliance with the norms and routines of formal schooling.
- A lesson gains in authenticity the more there is a connection to the larger social context within which students live. Two areas in which student work can exhibit some degree of connectedness are: (a) a real world public problem; i.e., students confront an actual contemporary or persistent issue or problem, such as applying statistical analysis in preparing a report to the city council on the homeless. (b) students' personal experiences; i.e., the lesson focuses directly or builds upon students' actual experiences or situations. High scores can be achieved when the lesson entails one or both of these.

## Appendix E: Scoring Criteria for Tasks

### General Rules

The main point here is to estimate the extent to which successful completion of the task requires the kind of cognitive work indicated by each of the three standards: Construction of Knowledge, Elaborated Communication, and Connections to Students' Lives. Each standard will be scored according to different rules, but the following apply to all three standards.

- If a task has different parts that imply different expectations (e.g., worksheet/short answer questions and a question asking for explanations of some conclusions), the **score should reflect the teacher's apparent dominant or overall expectations**. Overall expectations are indicated by the proportion of time or effort spent on different parts of the task and criteria for evaluation, if stated by the teacher.
- Take into account what students can reasonably be expected to do at the grade level.
- When it is difficult to decide between two scores, give the higher score only when a persuasive case can be made that the task meets minimal criteria for the higher score.
- If the specific wording of the criteria is not helpful in making judgments, base the score on the general intent or spirit of the standard described in the tips for scoring a particular AIW standard.

	Construction of Knowledge	Elaborated Communication	Connection to Students' Lives
<b>4</b>	N/A	Analysis / Persuasion / Theory. Explicit call for generalization AND support. The task requires explanations of generalizations, classifications and relationships relevant to a situation, problem, or theme, AND requires the student to substantiate them with examples, summaries, illustrations, details, or reasons. Examples include attempts to argue, convince or persuade and to develop and test hypotheses.	N/A
<b>3</b>	The task's <b>dominant expectation</b> is for students to interpret, analyze, synthesize, or evaluate information, rather than merely to reproduce information. To score high the task should call for interpretation of nuances of a topic that go deeper than surface exposure or familiarity.	Report / Summary. Call for generalization OR support. The task asks students <b>either</b> to draw conclusions <b>or</b> make generalizations or arguments, OR to offer examples, summaries, illustrations, details, or reasons, but not both.	The question, issue, or problem clearly resembles one that students have encountered or might encounter in their lives. The task explicitly asks students to connect the topic to experiences, observations, feelings, or situations significant in their lives.
<b>2</b>	There is some expectation for students to interpret, analyze, synthesize, or evaluate information, rather than merely to reproduce information.	Short-answer exercises. The task or its parts can be answered with only one or two sentences, clauses, or phrasal fragments that complete a thought.	The question, issue, or problem bears some resemblance to one that students have encountered or might encounter in their lives, but the connections are not immediately apparent. The task offers the opportunity for students to connect the topic to experiences, observations, feelings, or situations significant in their lives, but does not explicitly call for them to do so.
<b>1</b>	There is very little or no expectation for students to interpret, analyze, synthesize, or evaluate information. The dominant expectation is that students will merely reproduce information gained by reading, listening, or observing.	Fill-in-the-blank or multiple choice exercises.	The problem has virtually no resemblance to questions, issues, or problems that students have encountered or might encounter in their lives. The task offers very minimal or no opportunity for students to connect the topic to experiences, observations, feelings, or situations significant in their lives.

Appendix F: Scoring Tips for Task Rubric

<p><b>Tips for Scoring Construction of Knowledge</b></p>	<p><b>Tips for Scoring Elaborated Communication</b></p>	<p><b>Tips for Scoring Connection to Students' Lives</b></p>
<ul style="list-style-type: none"> <li>❑ The task asks students to organize and interpret information in addressing a concept, problem, or issue.</li> <li>❑ Consider the extent to which the task asks the student to organize, interpret, evaluate, or synthesize complex information, rather than to retrieve or to reproduce isolated fragments of knowledge or to repeatedly apply previously learned procedures. To score high the task should call for interpretation of nuances of a topic that go deeper than surface exposure or familiarity. Nuanced interpretation often requires students to read for subtext and make inferences. Possible indicators of interpretation may include (but are not limited to) tasks that ask students to consider alternative solutions, strategies, perspectives and points of view.</li> <li>❑ These indicators can be inferred either through explicit instructions from the teacher or through a task that cannot be successfully completed without students doing these things.</li> </ul>	<ul style="list-style-type: none"> <li>❑ The task asks students to elaborate on their understanding, explanations, or conclusions on important social studies concepts.</li> <li>❑ Consider the extent to which the task requires students to elaborate on their ideas and conclusions.</li> </ul>	<ul style="list-style-type: none"> <li>❑ The task asks students to address a concept, problem or issue that is similar to one that they have encountered or are likely to encounter in life outside of school.</li> <li>❑ Consider the extent to which the task presents students with a question, issue, or problem that they have actually encountered or are likely to encounter in their lives. Defending one's position on compulsory community service for students could qualify as a real world problem, but describing the origins of World War II generally would not.</li> <li>❑ Certain kinds of school knowledge may be considered valuable in social, civic, or vocational situations beyond the classroom (e.g., knowing how a bill becomes a law). However, task demands for "basic" knowledge will not be counted here unless the task requires applying such knowledge to a specific problem likely to be encountered beyond the classroom.</li> </ul>

## Appendix G: Email Correspondence Request for Tasks

Attn: 9-10th grade SS faculty

I recently had the opportunity to meet with each of you regarding a research study to determine what works to improve student learning outcomes in social studies. I appreciate your willingness to listen to my presentation and take part in this study. I have listed below the data that I'd like to collect from you this semester.

### Part I:

Please send me copies of three student assignments or assessments that you feel best indicate how well students understand your subject at a high level. So that I am clear on what you are asking your students to do, please include any materials necessary to help me understand the tasks and how they fit into the rest of your course. Please select tasks that relate to an instructional unit or a single lesson rather than midterm or final exams.

For each assignment, provide a general indication of when it will take place in your classroom this semester.

This information can be sent via email. Contact me if you have hard copies that you'd like me to pick-up.

### Part II:

I would like to observe a class that is associated with each task that you provide. I would like to observe the class that gives me the most insight into what students have done to prepare for each task. Once I get your tasks, I'll contact each of you to set up an observation schedule.

Please try to have the three student assignments or assessments to me no later than February 15. Coordinate with me if this is an unrealistic deadline for some reason. It is important to get this process started as soon as possible since we will undoubtedly be running into scheduling conflicts for observations (graduation exam, etc.). For those of you with interns, go ahead and provide the tasks this semester. Depending on your intern's teaching schedule, we may need to schedule the observations in the Fall.

Let me know if you have any questions. Thanks again for your participation in this study.

Lamont E. Maddox

## Appendix H: U.S. History Higher Order Assessment Resources

During the 1800s, the United States greatly expanded its territory through treaties, land purchases, and the use of force. Many Americans justified this expansion by saying that the United States had a “Manifest Destiny” to control all the land from the Atlantic to the Pacific. One of the main periods when the United States considered the idea of Manifest Destiny was during the Mexican-American War. The timeline below includes some important events in the relationship between the United States and Mexico during this time. Use it as a resource to assist you in thinking about the causes of the Mexican-American War and whether the U.S. was justified or wrong to declare war on Mexico.

### 1820s

---

- 1821 -Mexico gains independence from Spain
- 1823 -American citizens migrate to the Mexican territory of Texas to become Mexican citizens and obtain cheap land
- 1823 -Over the next few decades, Mexico’s government is unstable and weak as different groups fight for control

### 1830s

---

- 1835 -The Mexican central government attempts to exert more control over its territories (including Texas)
- 1835 -Texas declares independence from Mexico
- 1836 -Battle of the Alamo
- Mexican forces led by President Santa Anna are defeated at San Jacinto. Santa Anna signs a treaty granting Texas independence.
- 1836-1845 - Mexican Congress refuses to recognize the treaty
- U.S and some European nations recognize Texas independence

### 1840s

---

- 1845 -Texas becomes part of the United States. Mexican Ambassador leaves the United States in protest
- Mexico rejects Texas’ independence and its annexation by U.S.
- President Polk sends John Slidell to Mexico to try to purchase New Mexico and California and to address problems between the two nations. Mexican authorities refuse to meet with him
- Mar. 1846 -U.S. military forces enter territory claimed by both the United States and Mexico along the Rio Grande
- Apr. 1846 -Mexican forces cross the Rio Grande and enter the disputed territory. U.S. and Mexican forces clash
- May 1846 -U.S. declares war on Mexico
- 1846-1848 -Mexican-American War
- 1848 -U.S. wins the war. U. S. gains California, New Mexico, and other territories from Mexico as part of peace treaty

## Source Documents

**The following documents offer opposing views about Manifest Destiny and the Mexican-American war. Use the information from these documents and the timeline to decide whether the United States was justified in its war with Mexico.**

**Document 1:** (Boston Times, October 22, 1847)

The “conquest” [of Mexico] which carries peace into a land where military force is the usual basis for resolving conflict between competing groups, which establishes the reign of law where lawlessness has existed for a generation; which provides for the education and elevation of the great mass of the people, who have, for a period of 300 years been the slaves of an overbearing foreign race [the Spanish], and which causes religious liberty, and full freedom of mind to prevail where a [Catholic] priesthood has long been enabled to prevent all [other] religion, - such a “conquest” should be characterized as work worthy of a great people, of a people who are about to regenerate the world by asserting the supremacy of humans to decide their own fate [as opposed to their fate being decided by kings or dictators].

---

**Document 2:** Albert Gallatin, *Peace with Mexico* (New York, 1847, pp. 12-14.)  
Gallatin, a Swiss immigrant, served in a number of government positions including the House of Representatives and Secretary of the Treasury.

The people of the United States have been placed by God in a position never before enjoyed by any other nation. They are possessed of a most extensive territory, with a very fertile soil, a variety of climates, and a capacity of sustaining a population greater . . . than any other territory of the same size on the face of the globe. . . .

America’s mission is, to improve the state of the world, to be the “Model Republic,” to show that men are capable of governing themselves, and that this simple and natural form of government is the one that also makes the most people happy, is productive of the greatest development of the intellectual faculties, above all, the one that develops the highest standard of private and political virtue and morality.

In their foreign relations the United States, before this unfortunate war, always acted with justice. . . . The use of military force was always in self-defense. . . .

The allegation that the conquering of Mexico would be the means of enlightening the Mexicans, of improving their social state, and of increasing their happiness, is but the shallow attempt to disguise unbounded greed and ambition. Truth never was or can be spread by fire and sword, or by any other than purely moral means.

Documents excerpted and paraphrased from:

Rappaport, A. (1964). *The War with Mexico: Why did it Happen?* Berkeley: Rand McNally & Company.

## Appendix I: U.S. History Higher Order Assessment Instructions

**Part I:** Assume the role of a concerned citizen in 1847. The war with Mexico is nearing the end of its first year. U. S. newspapers are full of commentaries about the war. You have decided to write an editorial for one of these newspapers. Using information from the timeline, the source documents, and your knowledge of the time period, write a persuasive essay that takes a position on whether Manifest Destiny adequately justifies going to war with Mexico. Specifically, **is using Manifest Destiny to justify war [in Mexico] a violation of American ideals [and therefore wrong] or does pursuing Manifest Destiny in Mexico ultimately promote the greater good?**

Your editorial should meet the following guidelines:

### *Requirements:*

1. Your editorial must include a minimum of 4 paragraphs as described below.
2. Your editorial should use persuasive language and should be written to the readers of the newspaper.
3. Your editorial should be written in the 1<sup>st</sup> person, plural tense – “we”. For example: I believe the war is just because.... **Or** We took the wrong approach because.....
4. Note: The format provided below is meant to be used as an outline for writing your editorial. Your final paper should be written as one coherent, continuous essay. However, to assist in grading, please identify each part of your editorial by using the section headers provided below (i.e. Section I, Section II, etc.).

### *Editorial Format:*

#### **Section I: Introduction**

Briefly describe the situation between the United States and Mexico. Discuss the most important events that contributed to the war. Use this information to lead to a final statement that clearly describes your position on the war as it relates to Manifest Destiny (For example: The United States is wrong to use Manifest Destiny to go to war with Mexico **Or** The United States has every right to pursue its Manifest Destiny by conquering Mexico).

#### **Section II: Support your argument**

Provide at least two or three distinct reasons to defend your position. Your reasons should be supported by evidence from the timeline, the source documents, and your knowledge of the time period. Make sure your arguments in this paragraph clearly relate to Manifest Destiny.

#### **Section III: Address the arguments of those who disagree with you.**

Acknowledge the arguments of those who might take an opposing position on this issue. In doing so, provide two or three distinct reasons your opponents might use to disagree with your point of view. Cite information from the timeline, the



source documents, and your knowledge of the time period to support this perspective.

**Section IV: Conclusion**

Respond to the arguments of your opponents and summarize your most persuasive points.

**Part II:**

Step out of the role of being a citizen of the 1840s and answer the following question based on your own opinion.

Consider the role of the United States in world affairs today. Does America still have a special destiny or mission in the world? If so, what is it and how should it be accomplished? If not, explain why you think it does not.

## Appendix J: Advanced Placement Higher Order Assessment Student Resources

During the 19<sup>th</sup> century, many liberal nationalists in Germany sought to organize the separate German states into one nation-state based on principles of representative government. The “German Question” of how to define the boundaries of the new *Reich* was one of many problems that made unity and freedom for the German people difficult to accomplish. The timeline below includes some important events that will help answer the question: **Should the unification of all Germanic peoples within one nation be endorsed (supported) by the German people and encouraged by other nations in 1870?**

### Early 1800s

---

- |        |  |
|--------|--|
| 1814/5 | -German Confederation born at Congress of Vienna |
| 1834   | -Custom Union called the Zollverein established  |

### 1840s

---

- |           |   |
|-----------|---|
| Feb. 1848 | -Revolution in France; overthrow of the monarchy of King Louis-Philippe; proclamation of the creation of the French Second Republic |
| Mar. 1848 | -Uprisings in some German states; granting of constitutional reforms in Prussia   |
| 1848/1849 | -Revolutions in Italy, Vienna, Budapest, and Prague   |
| May 1848  | -Frankfurt Assembly meets and proposes a plan for the unification of Germany; Prussian king refuses to take the crown               |

### 1860s

---

- |      |   |
|------|---|
| 1862 | -Bismarck becomes prime minister of Prussia   |
| 1862 | -Bismarck gives the “Blood and Iron” speech to the Budget Committee of Prussia’s lower parliamentary house  |
| 1864 | -Danish-Prussian War  |
| 1866 | -Austro-Prussian War  |
| 1867 | -North German Federation formed.<br>-The constitution of the North German Confederation serves as a model for that of the German Empire, with which it merged in 1871 |

### 1870s

---

- |           |   |
|-----------|---|
| June 1870 | -Controversy involving the Hohenzollern candidacy for the Spanish throne                        |
| July 1870 | -Bismarck publishes the edited Ems dispatch   |
| July 1870 | -Franco-Prussian War  |
| Jan. 1871 | -Proclamation of the German Empire at Versailles  |
| May 1871  | -Treaty of Frankfurt ratified between France and Germany<br>Germany annexes Alsace and Lorraine |

## Source Documents

**The following documents offer additional information that will help you address the assessment question.**

**Document 1:** August Bebel Criticizes the Franco-Prussian War and the Annexation of Alsace-Lorraine in a Speech before the North German Reichstag (November 26, 1870)

...In my opinion, the *principle of nationality* is a thoroughly *reactionary* principle. You will admit that if we were to apply the principle of nationality in its pure form in Europe, there would be no end in sight to war; the peoples' mission would always and exclusively be to make war, to work only to make war possible. On the basis of the principle of nationality we would have to cede [give away] Poland, return northern Schleswig, get rid of South Tyrol and Trento, and relinquish many Slavic-speaking regions; on the other hand, we would have to annex [make part of Germany] parts of Switzerland, the Netherlands, and Belgium. As I have already mentioned, according to the principle of nationality, we would not be able to get out of war. The peoples would tear each other apart until the end of time. Nationality means but little; in my view, it has merely a secondary importance for the political life of a state. The highest and most fundamental idea in the political life of a state must be the internal satisfaction of peoples through their institutions, their right to self-determination.

Translation: Erwin Fink

“August Bebel. *Sein Leben in Dokumenten, Reden und Schriften*”, a document by Helmut Hirsch. In *Forging an Empire: Bismarckian Germany, 1866-1890*, edited by James Retallack, volume 4, *German History in Documents and Images*, German Historical Institute, Washington, DC ([www.germanhistorydocs.ghi-dc.org](http://www.germanhistorydocs.ghi-dc.org)).

**Document 2:** From the Debates in the German National Assembly on the territories to be included as part of a German nation-state – Little Germany or Greater Germany? – 1848-9

Context: The German National Assembly is debating the following alternatives for unification:

"Lesser German Solution": A united Germany, led by Prussia, without Austria.

"Greater German Solution": A German state that includes most of the German speaking population of Europe. Some wanted the German speaking territories of the Austrian Empire. Others favored including all of Austria as part of the German nation-state.

Venedy (Representative for Cologne):

‘...We have come here, Gentlemen, to constitute Germany’s Unity, and we are met with the proposal that we throw a part of Germany out of Germany. On that day when we only discuss this proposition, we will be discussing the division of Germany. The German nation, Gentlemen, has already suffered enough, but she has finally prevailed and has sent us here to constitute Germany, and they want us to sell off a part of Germany. I have come here...with the firm decision to stand or to fall with the assembly.

But I do not want to sit here a moment longer if Austria is not here too [as a member of the new German empire].

Moritz Mohl (Representative for Stuttgart):

‘...We are 40 million Germans; we do not need to fear these scattered little nations. There are perhaps five million Czechs: there are not five million Magyars, still fewer Croatians and even fewer Wallachians etc....All these nations [within Austria] can do no disadvantage to German nationality; it is however of the very greatest importance that they combine with Germany, and that with Germany they form a *Reich* of seventy million persons. Gentlemen! I ask you, when these seventy million people are represented in a German parliament, when this parliament through its influence nominates the ministers of this great *Reich*, and when nothing occurs to the disadvantage of this great *Reich* of seventy million people; I ask you, which power in Europe, even Russia with her sixty-six millions, or France with her thirty-six millions, which power in Europe will be powerful enough to challenge this great *Reich*? I ask, whether this German *Reich* is then not in a condition to dictate war and peace to the whole world; I ask you, to consider this.... Gentlemen! This thought about the entry of the whole of Austria within the German Federal State; I beg you to fix your eyes on this thought, telling yourselves that it removes every difficulty....

## Appendix K: Advanced Placement Higher Order Assessment Instructions

**Part I:** In 1870, Germany successfully defeated France in the Franco-Prussian War. One outcome of this war was a renewed effort to create a unified nation-state for all Germans. The topic of unification was discussed in parliamentary proceedings, newspapers, and official correspondence between statesmen. The debate over German unification raised broader issues of how a nation’s boundaries should be drawn.

Assume the role of a German citizen in 1870. You have decided to write an editorial for a German newspaper on the issue of unification. Remember, it is 1870 and the ultimate solution of 1871 has not yet been decided. The editorial should reflect your judgment of what the best solution *should be*. Using information from the timeline, the source documents, and your knowledge of the time period, write a persuasive essay that takes a position on the following question: **Should the unification of all Germanic peoples within one nation be endorsed (supported) by the German people? Would other nations likely support it?** In framing your response, consider a number of factors such as the potential military, political, social, and economic consequences of unification.

Your editorial should meet the following guidelines:

### *Requirements:*

1. Your editorial must include a minimum of 4 paragraphs as described below.
2. Your editorial should use persuasive language and should be written to the readers of the newspaper.
3. Your editorial should be written in the 1<sup>st</sup> person, plural tense – “we”. For example: I believe other nations should encourage unification because... or We should not support unification because...
4. Note: The format provided below is meant to be used as an outline for writing your editorial. Your final paper should be written as one coherent, continuous essay. However, to assist in grading, please identify each part of your editorial by using the section headers provided below (i.e. Section I, Section II, etc.).

### *Editorial Format:*

#### **Section I: Introduction**

Briefly describe the most significant events in the road to German unification up to this point (1870). In doing so, remember to assume the perspective of a citizen of this time period who is not aware of the actual events to come regarding the unification of Germany. Use this information to lead to a final statement that clearly describes your position on German unification (For example: The unification of all Germanic peoples within one nation should be endorsed by the German people because...or Germans should oppose unification of all Germanic peoples within one nation because ...).

**Section II: Support your argument**

Provide at least two or three distinct reasons to defend your position. Your reasons should be supported by evidence from the timeline, the source documents, and your knowledge of the time period.

**Section III: Address the arguments of those who disagree with you**

Acknowledge the arguments of those who might take an opposing position on this issue. In doing so, provide two or three distinct reasons your opponents might use to disagree with your point of view. Cite information from the timeline, the source documents, and your knowledge of the time period to support this perspective.

**Section IV: Conclusion**

Respond to the arguments of your opponents and summarize your most persuasive points.

**Part II:** Step out of the role of being a citizen of the 1870s and answer the following question based on your own opinion.

Some policy makers today support the formation of nation-states based on common ethnic, cultural, or religious identities as a way to stop violence in regions around the world (i.e. a Palestinian state; separate Kurd, Sunni, & Shiite states instead of a united Iraq). To what extent, if any, should the U.S. support the ambitions of ethnic, cultural, or religious groups seeking to secure their own nation-states today? In your response, consider the pros and cons of supporting these sorts of new nation-states and discuss why the course of action you recommend is preferable to the position taken by those opposing your view.

## Appendix L: Proctor Instructions

Ensure your students have several blank sheets of paper available and a writing utensil.

Step 1: You should have a set of notecards (by block) that includes the names of your students and their corresponding student numbers. Pass the notecards out to your students and allow them to transfer their student number to their answer sheets. Students should write their number on each page they intend to turn in. They should not write their name on the assessment – just their student number.

Step 2: Read to students:

Today you will be writing an essay that measures your ability to think critically about an issue of historic and contemporary importance. This assessment is being tested by \_\_\_\_\_ Schools as an additional way for students to demonstrate what they've learned in their social studies courses. Do your very best since the assessment will be used as one indication of how well you can apply your knowledge of European history. **[insert how the assessment will be graded for your individual class]**

This assessment is primarily a test of your ability to reason and make persuasive arguments related to the forming of nation-states and, in particular, German unification. It includes two main parts. The first part requires you to write an editorial. A timeline and two historical documents are provided to help you with this task. The second part asks you to state and support your opinion about American support for new nation-states today. **Partial credit is awarded, so it is in your best interest to attempt to answer each part of the assessment.** You can still earn points even if you do not have a great deal of prior knowledge about the topics included on this test.

You may underline passages or take notes on the materials provided to you for this assessment. However, your final response should be provided on separate sheets of paper. All testing materials will be turned in by the end of the testing period. When you finish the assessment, turn over your work and wait for your teacher to come by and pick it up. Please remain quiet throughout the testing period so your peers can concentrate.

As the proctor for this essay, I cannot provide any hints, answers, or suggestions to you as you take this exam. I can restate the directions if you don't understand what you are being asked to do.

You have 1 hour to complete this essay. What are your questions?

## Appendix M: Scoring Rubric for Advanced Placement Higher Order Editorial

### Part I

1. **Position Statement.** Does the student take a clear position on the question? (Y=1, N=0). This 1-2 sentence statement can be found anywhere in the essay. In order to take a clear position, the student's statement must specifically indicate what the German people would be endorsing (i.e. a unified state for *all* German-speaking people, no unification at all, a limited German state with Prussian leadership, etc.).
2. **Historical Context of Problem.** How well is the problem defined in paragraph one before the student provides arguments related to the focus question of the editorial? Does the student appear to understand the events and/or historic forces (i.e. liberalism, nationalism) related to unification described in the opening paragraph?

0= No background context provided.

Assign a 0 when:

- There is a position statement with no other information
- The introduction includes vague statements with no real factual information from the timeline
- The introduction just includes persuasive arguments instead of background information
- The paragraph includes more inaccurate statements than valid contextual information

1= Some background context provided. The student provides a brief mention of some historical events over the course of 1-2 sentences as part of the introductory paragraph.

Scoring notes:

- Information that is copied from the top of the timeline does not count.
- A level one score is characterized by very limited information in paragraph 1 that closely follows the timeline. However, a student might also get a 1 score if he/she copies virtually every event from the timeline (almost verbatim).
- Do not assign a "1" if the student only mentions the war Prussia just won against France.

2= Historical context is well defined. The student provides a clear and coherent introduction to the editorial. Historical events are introduced strategically to build up to a thesis statement. The paragraph includes at least two sentences of relevant historical information or a particularly strong description of the problem.

Scoring notes:

- The paragraph must stick to the appropriate time period (1870)
- Inaccurate statements will drop the score to a 1.
- Look for originality in how the information is introduced to the reader.



- A “2” should be strongly considered when students incorporate appropriate ideas/topics/events not listed on the timeline.

3. **Persuasiveness.** To what extent does the essay demonstrate persuasive reasoning? Read the entire essay to evaluate the persuasiveness standard. The underlined portion of the standard is the main factor to consider in assigning a score.

0= Unsatisfactory. The student has failed to take a stand on the question, or has taken a stand, but has failed to provide a single persuasive reason. The response may indicate that the student didn’t fully understand the question. Overall, the response has no chance of persuading the reader.

1 or 2= Minimal. The student has taken a stand on the question (which may be flawed) and provided at least one persuasive reason to back up this stance. Faulty assumptions, undermining, or irrelevant reasons could result in an unsatisfactory score if they reduce the persuasiveness of the argument. Overall, however, the response is unlikely to persuade the reader.

1= The student provided a single persuasive reason to support his/her argument. The reason *may* have no clear connection to the question of how a unified Germany should be created (focusing instead on the desirability of unification). For example, the student might argue that unification is a good thing without ever describing what type of unified German state the people should endorse.

2= This score is assigned when a student provides multiple arguments that focus entirely on the pros or cons of unification (demonstrating a flawed understanding of the question). It might also be awarded when a student provides a single persuasive argument that is well stated (typically requiring more than one sentence), but not described at the level of detail needed for a level 3 score. The argument, by itself, remains unlikely to persuade the reader.

3= Adequate. The student has taken a stand on the question and has provided two or more persuasive reasons. The arguments in the essay have a clear relation to the question. Elaboration of reasons is not necessary here. The presentation of only one persuasive reason can result in a score of “adequate” if useful elaboration is included. Undermining reasons, faulty assumptions, or irrelevant reasons can possibly reduce the score to a 2. Overall, the response has a chance of persuading the reader.

\*When trying to determine if a single persuasive reason is thorough enough for an adequate score, consider the main criteria for this standard. Did the student’s elaboration result in an overall argument that has a chance of persuading the reader?

4=Elaborated. The student has taken a stand, provided two or more persuasive reasons, and has provided elaboration on at least one of those reasons (i.e. accurately referencing documents, providing examples, etc.). Presentation of many persuasive reasons (at least three) can also produce this score. Overall, the response is likely to persuade the reader.

\*The student must address in some way the potential reaction of other countries to German unification (the 2<sup>nd</sup> part of the focus question) to get a 4.

5= Exemplary. The student's response meets criteria for "elaborated", and demonstrates (a) at least two elaborated persuasive reasons, and (b) an argument so clear and coherent (i.e. no significant undermining reasons, faulty assumptions, or irrelevant reasons) and grammatically correct as to merit public display as an outstanding accomplishment for a high school student. Overall, the response is more likely to persuade the reader than the elaborated response.

#### **4. Low Level Dialectical Reasoning.** To what extent are opposing arguments recognized and developed?

0 = opposing arguments are not addressed in the editorial or they are not described fully enough to make sense. A student might also receive this score if the opposing arguments are not accurate.

\*Scoring tip: "Described fully enough to make sense" – take this literally to mean that you can't understand what the student is saying. If you can reasonably understand the point being made by the student, and it represents an accurate opposing view, assign a 1.

1 = includes one argument that accurately represents an opposing viewpoint on the issue. The argument is described in *minimal* detail with little or no use of historical evidence. Strong opposing perspectives may be ignored or greatly simplified so they can be easily refuted later in the essay.

2= includes multiple arguments of the sort described for a level 1 score.

3 = includes at least one well developed argument that accurately represents an opposing viewpoint on the issue. The student seems to understand the opposing argument(s) he/she is representing. The degree of development in the paragraph suggests the student gave more than cursory consideration to the opposing perspective.

Scoring tip:

\*The likelihood of a level 3 score increases when a student dedicates an entire paragraph to explaining opposing views instead of following the pattern of presenting an argument only to immediately shoot it

down

\*Look for use of the documents to back up the opposing perspective

\*Look for an explanation of the opposing view that covers at least a couple of consecutive sentences.

5. **Quality of Final Position.** How well does the student synthesize opposing viewpoints and offer persuasive counter-arguments to arrive at a well supported final position?

0=Unsatisfactory. The student doesn't provide a conclusion (although some opposing arguments might be addressed in section 3) or the conclusion is very brief.

Assign a 0 when:

- No concluding paragraph is provided
- The student doesn't mention/restate any key points
- The concluding paragraph mostly includes inaccurate/vague statements
- The paragraph mainly quotes (perhaps without citing) directly from a source document with no elaboration on the part of the student
- The concluding paragraph mostly includes arguments based on a future Germany that doesn't exist in 1870
- The concluding paragraph actually reduces the overall persuasiveness of the editorial based on the presence of random, unintelligible, or inaccurate statements.

1= Adequate. The student provides a concluding paragraph that incorporates a response to critics (perhaps at the end of the previous paragraph) and *brief mention* of at least 1 key point made in the editorial. The final position generally does not add to the persuasiveness of the essay (perhaps because it is overly brief, vague, ignores major holes in argumentation, etc.). Some significant questions may be left unresolved for the reader.

Scoring tip:

\*Simply responding to critics is not enough. The student must conclude the paragraph by listing or mentioning 1-2 key points - perhaps in conjunction with a restatement of the thesis (this could be 1 sentence).

2= Approaching Satisfactory. The basic standards for a level 1 score are met. A level 2 paragraph features a stronger summary of the key points made in the essay *or* a more persuasive response to the views of opponents. Overall, the paragraph adds to the persuasiveness of the essay, but there is little evidence the student genuinely weighed the views of critics when crafting his/her final

position (no higher level dialectical reasoning). Undermining reasons, faulty assumptions, or irrelevant reasons can possibly reduce the score to “adequate”.

Scoring tip:

“stronger summary of the key points” – A good solid paragraph (3 sentence minimum) that clearly articulates the students’ point of view.

3= Satisfactory. The student synthesizes the views of opponents (perhaps at the end of section III) and takes these arguments into account when developing a persuasive final position. The final position includes at least 2-3 key points.

Scoring tips:

\*Look for tight argumentation (not many unanswered questions), passionate language, a thoughtful response to critics, and reinforcement of key points/ideas.

\*Look for language that suggests the student really considered the opposing view (i.e. my opponents make a good point when they say \_\_\_\_\_, but I feel they are overlooking....; I concede that German unification might cause \_\_\_\_\_, but I wonder if my critics have considered....)

## Part II

1. **Decision-making.** Does the student take a clear position regarding whether the U.S. should support the formation of new nation-states based on common traits? (Y=1; N=0).
2. **Persuasiveness.** To what extent does the response demonstrate persuasive reasoning? Read the entire essay to evaluate the persuasiveness standard. The underlined portion of the standard is the main factor to consider in assigning a score.

0=Unsatisfactory. The student has failed to take a stand on the question, or has taken a stand, but has failed to provide a single persuasive reason. The response may indicate that the student didn’t fully understand the question. Overall, the response has no chance of persuading the reader.

1=Minimal. The student has taken a stand on the question and provided at least one persuasive reason to back up this stance. Faulty assumptions, undermining, or irrelevant reasons could result in an unsatisfactory score if they reduce the persuasiveness of the argument. Overall, however, the response is unlikely to persuade the reader.

2= Adequate. The student has taken a stand on the question and has provided two or more persuasive reasons. Elaboration of reasons is not necessary here. The

presentation of only one persuasive reason can result in a score of “adequate” if useful elaboration is included. Undermining reasons, faulty assumptions, or irrelevant reasons can possibly reduce the score to a 2. Overall, the response has a chance of persuading the reader.

3=Elaborated. The student has taken a stand, provided two or more persuasive reasons, and has provided elaboration on at least one of those reasons (i.e. providing examples, etc.). Presentation of many persuasive reasons (at least three) can also produce this score. Overall, the response is likely to persuade the reader.

4= Exemplary. The student’s response meets criteria for “elaborated”, and demonstrates (a) at least two elaborated persuasive reasons, and (b) an argument so clear and coherent (i.e. no significant undermining reasons, faulty assumptions, or irrelevant reasons) and grammatically correct as to merit public display as an outstanding accomplishment for a high school student. Overall, the response is more likely to persuade the reader than the elaborated response.

## Appendix N: Scoring Rubric for Manifest Destiny Higher Order Assignment

### Part I

1. **Position Statement.** Does the student take a clear position on the question? (Y=1, N=0). This statement can be found anywhere in the essay. The position must relate to Manifest Destiny.
2. **Historical Context of Problem.** Does the student appear to understand the events that contributed to the Mexican-American War? How well is the problem defined in paragraph one before the student engages in arguments for or against America's actions?

0= No background context provided.

Assign a 0 when:

- There is a thesis with no other information
- The introduction includes vague statements with no real factual information
- The introduction just includes persuasive arguments instead of background information
- The paragraph includes more inaccuracies or seemingly random statements than valid contextual information

1= Some background context provided. The student provides some historical context for the essay (at least one historical event). The event should serve as context, not as part of an argument.

- Information that is copied from the top of the timeline sheet (i.e. the first sentence, the definition of Manifest Destiny) does not count.
- Events may be listed (perhaps verbatim) from the timeline. It may be unclear whether the student truly understands the problem – especially if some of the events are inaccurately stated.

2= Historical context is well defined. The student demonstrates an understanding of the problem and uses at least some language that differs from the source documents.

Key indicators:

- The paragraph sticks to the appropriate time period (1847). The student should not state that the U.S. gained New Mexico, California, and Texas as a result of the war.
- The paragraph suggests some elaborated understanding beyond what is on the timeline (see scoring tips).
- The paragraph is generally free of inaccurate statements.
- The paragraph should include or at least reference key events that immediately led to the Mexican-American War (border dispute, annexation of Texas). The historical context is not well defined if the

student exclusively talks about the war for Texas' independence that happened over ten years before the decision point of this essay.

3. **Persuasiveness.** To what extent does the essay demonstrate persuasive reasoning? Does the student relate his/her response to Manifest Destiny? Read the entire essay to evaluate the persuasiveness standard. The underlined portion of the standard is the main factor to consider in assigning a score.

0= Unsatisfactory. The student has failed to take a stand on the question, or has taken a stand, but has failed to provide a single persuasive reason. The response may indicate that the student didn't fully understand the question. Overall, the response has no chance of persuading the reader.

1 or 2= Minimal. The student has taken a stand on the question and provided at least one persuasive reason to back up this stance. The "stand" may be focused entirely on whether the war was right or wrong with no reference to Manifest Destiny. Faulty assumptions, undermining, or irrelevant reasons could possibly reduce the score from a 2 to a 1 or from minimal to unsatisfactory. Overall, however, the response is unlikely to persuade the reader.

1=The student provided a single persuasive reason to support his/her argument. The reason may have no clear connection to Manifest Destiny.

Examples of arguments without a connection to Manifest Destiny:  
The U.S. was acting in self-defense because Mexico attacked first  
The war was justified because Mexico refused to meet with Slidell  
The U.S. needed the land for a growing population  
Stealing land is wrong (assuming the student didn't connect this statement to Manifest Destiny)

2=The student's essay mainly focused on whether the war was right or wrong. In this context, the student provided multiple persuasive reasons, or a single persuasive reason described in greater depth (several sentences), to support his/her position.

\*Note: A single persuasive argument with a clear connection to manifest destiny that does not contain enough elaboration for a "3" would also receive this score.

3=Adequate. The student has taken a stand on the question and has provided two or more persuasive reasons. The arguments in the essay have a clear relation to Manifest Destiny. Elaboration of reasons is not necessary here. The presentation of only one persuasive reason can result in a score of "adequate" if useful elaboration is included. Undermining reasons, faulty assumptions, or irrelevant reasons can possibly reduce the score to "minimal." Overall, the response has a chance of persuading the reader.

\*When trying to determine if a single persuasive reason is thorough enough for an adequate score, consider the main criteria for this standard. Did the student's elaboration result in an overall argument that has a chance of persuading the reader?

4=Elaborated. The student has taken a stand, provided two or more persuasive reasons, and has provided elaboration on at least one of those reasons (i.e. accurately referencing documents, providing examples, etc.). Presentation of many persuasive reasons (at least three) can also produce this score. The arguments presented by the student have a clear connection to Manifest Destiny. Overall, the response is likely to persuade the reader.

5=Exemplary. The student's response meets criteria for "elaborated", and demonstrates (a) at least two elaborated persuasive reasons, and (b) an argument so clear and coherent (i.e. no significant undermining reasons, faulty assumptions, or irrelevant reasons) and grammatically correct as to merit public display as an outstanding accomplishment for a high school student. Overall, the response is more likely to persuade the reader than the elaborated response.

4. **Low Level Dialectical Reasoning.** To what extent are opposing arguments developed?

0= opposing arguments are not addressed in the editorial or they are not described fully enough to make sense. A student might also receive this score if the opposing arguments are not accurate.

1= includes one argument that accurately represents an opposing viewpoint on the issue. The argument is described in minimal detail with little or no use of historical evidence. Strong opposing perspectives may be ignored or greatly simplified so they can be easily refuted later in the essay.

2= includes multiple arguments of the type described for a level 1 score.

3= includes at least one well developed argument that accurately represents an opposing viewpoint on the issue. The student seems to understand the opposing argument(s) he/she is representing. The degree of development in the paragraph suggests the student gave more than cursory consideration to the opposing perspective.

5. **Quality of Final Position.** How well does the student synthesize opposing viewpoints and offer persuasive counter-arguments to arrive at a well supported final position?

0=Unsatisfactory. The student doesn't provide a conclusion or the conclusion consists of a single sentence that restates the student's opinion.

Assign a 0 when:

- No concluding paragraph is provided



- The student doesn't make/restate any arguments
- The concluding paragraph mostly includes inaccurate/vague statements
- The paragraph mainly quotes (perhaps without citing) directly from a source document with no elaboration on the part of the student
- The concluding paragraph mostly includes arguments based on a future America that doesn't exist in 1847
- The concluding paragraph actually reduces the overall persuasiveness of the editorial based on the presence of random, unintelligible, or inaccurate statements.

1=Adequate. The student lists or mentions 1-2 key points made in the essay. This may be in conjunction with a restatement of the thesis. The conclusion generally does not add to the persuasiveness of the essay and the arguments of critics are given little to no consideration. Some significant questions may be left unresolved for the reader.

2=Approaching Satisfactory. The student summarizes 1-2 key points made in the essay or adds a final persuasive reason that is new. Overall, the paragraph adds to the persuasiveness of the essay, but there is little evidence the student genuinely weighed the views of critics when crafting his/her final position. Undermining reasons, faulty assumptions, or irrelevant reasons can possibly reduce the score to "adequate".

3=Satisfactory. The student synthesizes the views of opponents (perhaps at the end of section III) and takes these arguments into account when developing a persuasive final position. The final position should include at least 2-3 key points.  
 "takes these arguments into account" = mentions or references them in final paragraph

**Part II** assesses the connectedness to the real world standard. Essays are evaluated based on the extent to which the student connects the disciplinary topic of Manifest Destiny to contemporary issues or events that have personal relevance in their own life.

2= Explicit Connection. There is an explicit connection being made between classroom knowledge (Manifest Destiny) and contemporary situations outside the classroom.

1=Possible Connection. The student's response hits on themes associated with Manifest Destiny/American Exceptionalism, but this may not have been intentional on the part of the student. The student might also receive this score if he/she makes valid historical references demonstrating how a particular mission might be exemplified across time.

0= No connection. It isn't clear whether the student recognizes any parallels between Manifest Destiny in the 1800s and U.S. actions today.

Indicators: a totally off topic response, vague responses (world peace), etc.

### **Scoring Tips:**

#### Position Statement:

- Does the student clearly weigh in on one side of the issue (no waffling)?
- Does the student's position clearly relate to Manifest Destiny? Simply stating that the war was wrong or right does not count.
- Look for an actual statement. In some cases, you will be able to infer the student's position based on arguments throughout the essay. However, a 1 is only assigned for a concise statement (1-2 sentences) that clearly indicates the students' position on the idea of Manifest Destiny as it pertains to the Mexican-American War.
- It is possible for a student to take a stand on the question without having a clear position statement. The persuasiveness score can still be high assuming the student's position on Manifest Destiny can reasonably be inferred.

#### Historical Context:

- A great deal of variation can exist at level 2. A student with one background event and a student with an entire page of events can potentially get the same score.
- A paragraph that is not set within the United States (i.e. student assumes perspective of a Mexican citizen) can receive a 2 if background events are still explained accurately.
- Close adherence to the timeline is an indicator that the student might not possess much depth of knowledge on the topic (i.e. inclusion of virtually every event on the timeline in paragraph 1 or misreading the timeline to suggest that Texas gained its independence as a result of the Alamo, etc.).
- Indicators of more in-depth understanding that would support a score of 2:
  - purposely selecting key events rather than trying to cover every topic listed on the timeline.
  - properly using the phrase "Manifest Destiny"
  - incorporating appropriate ideas/topics/events not listed on the timeline
  - using style and language that differs substantially from the timeline

#### Persuasiveness:

Any score above the minimal level requires a connection to be made to Manifest Destiny. The connection doesn't necessarily have to be explicit if you can reasonably infer that the student understands Manifest Destiny and that his/her arguments closely fit the question. A particularly strong conclusion can add to the persuasiveness score.

Quality of final position:

In judging between a 1 or 2, consider how developed the paragraph is (summarize vs. mention) and its persuasiveness. A weak concluding argument (in terms of logic) would likely receive a 1.

Appendix O: Authentic Pedagogy Scores

**Minimal Authentic Pedagogy**

Roy's Authentic Pedagogy scores

Task Name	Task Score (3-10)	Instruction Score (4-20)	Final Authentic Pedagogy (7-30)
Political Cartoon	4	5	
Industrial Revolution Illustrated Timeline	6	6	
Teach a Lesson	4	4	
Average Task/Instruction Scores	4.6	5	9.6

Andy's Authentic Pedagogy scores

Task Name	Task Score (3-10)	Instruction Score (4-20)	Final Authentic Pedagogy (7-30)
U.S. History Project	5	5	
Reformers of the 1800s	5	7	
Manifest Destiny Questions	4	7	
Average Task/Instruction Scores	4.6	6.3	10.9

Jason's Authentic Pedagogy scores

Task Name	Task Score (3-10)	Instruction Score (4-20)	Final Authentic Pedagogy (7-30)
Presidential Research	8	4	
Declaration Activity	7	5	
Daily Life of Civil War Soldiers	5	6	
Average Task/Instruction Scores	6.6	5	11.6

### Limited Authentic Pedagogy

#### Amy's Authentic Pedagogy scores

Task Name	Task Score (3-10)	Instruction Score (4-20)	Final Authentic Pedagogy (7-30)
Absolute Monarchy of your Own	4	8	
Ideal Form of Government Debate	9	10	
Renaissance Ball	4	4	
Average Task/Instruction Scores	5.6	7.3	12.9

#### Phillip's Authentic Pedagogy scores

Task Name	Task Score (3-10)	Instruction Score (4-20)	Final Authentic Pedagogy (7-30)
Washington's Farewell Address	8	8	
Reformers Lesson	7	4	
Manifest Destiny Painting Analysis	6	7	
Average Task/Instruction Scores	7	6.3	13.3

### Moderate Authentic Pedagogy

#### Lauren's Authentic Pedagogy scores

Task Name	Task Score (3-10)	Instruction Score (4-20)	Final Authentic Pedagogy (7-30)
Industrial Revolution Documentary	8	8	
PR Campaign Billboard Assignment	8	14	
French Revolution Storybook	5	N/A	
Average Task/Instruction Scores	7	11	18

#### Ryan's Authentic Pedagogy scores

Task Name	Task Score (3-10)	Instruction Score (4-20)	Final Authentic Pedagogy (7-30)
Czar Nicholas Think Aloud	8	14	
Political Cartoon Analysis	8	12	
Me Card	7	14	
Average Task/Instruction Scores	7.6	13.3	20.9

Lee's Authentic Pedagogy scores

Task Name	Task Score (3-10)	Instruction Score (4-20)	Final Authentic Pedagogy (7-30)
Industrial Revolution Editorial	8	15	
Truman Think Aloud	8	11	
Industrial Revolution Illustrated Timeline	7	15	
Average Task/Instruction Scores	7.6	13.6	21.2

Appendix P: Manifest Destiny Painting

American Progress by John Gast







Appendix Q: WWII Political Cartoon



Appendix R: Moderate Authentic Pedagogy Task

Truman Considers the Berlin Crisis  
Is the U.S. justified in imposing its will in Europe?

Instructions for the Truman Decision-making groups:

In a meeting on this crisis, you will hear from George Marshall (Sec. of State), George Kennan (ambassador to the U.S.S.R.), Henry Wallace (former Sec. of Commerce), and Walter Lippman (well known journalist). Listen carefully to each of their positions and recommendations. Record these and any concerns you have below. After hearing from each person, discuss the options available to President Truman. Brainstorm the strengths and weaknesses, benefits and dangers of each position.

Advisor	Recommendation	Concerns	Strength	Weakness
Marshall				
Kennan				
Wallace				
Lippman				

**In coming to your decision consider:**

1. What are the strongest arguments to be made for each option?
2. What are the strongest arguments against each option?
3. Is the U.S. justified in imposing its will in Europe? If so, how?
4. Is the U.S. justified in withdrawing from the conflict?
5. Does our moral responsibility to a people cut off by an outside force outweigh all other political/practical alternatives?
6. What decision will bring about the best solution for the U.S.? Europe? The world?

You will justify your decision to the American people in a speech – be sure to address each of these considerations as you plan your thoughts (collectively).

**What course of action did your Truman group choose? Justify your choice! (use other paper as needed)**



## Appendix S: Content Analysis Explanation and Examples

Several publications produced by the Alabama Department of Education provide information about the social studies graduation exam (Alabama Department of Education, 2009b; Morton, 2009; Richardson, 2000). Alabama began minimum competency testing in 1977. Three editions of the graduation exam have been created since, with the latest implemented in 1998 (Morton, 2009, p. 11). The third edition was the first to include a social studies subtest. The social studies graduation exam has 100 questions. Students must answer 54 correctly to pass.

The social studies graduation exam associated with this study went into effect with the class of 2003. In order to receive a diploma, students must pass all of the graduation exams. However, students can get an alternative diploma called the Alabama High School Diploma with Credit-Based Endorsement if they pass reading, math, and one other graduation exam. Students initially take the social studies exam in the tenth grade for practice. This test counts if it is passed. Students get four additional attempts to pass the exam before the end of their senior year. After graduation, exited students can take the exam as many times as they want during regularly scheduled times (Morton, 2009, p. 4).

No previous versions of the social studies graduation exam have been released and little information was available regarding the process for determining cut scores. The best source of information on the test items came from Bulletin 2000, No. 49: a publication provided to the general public by the Alabama Department of Education. This publication was designed to enable students to understand the general format of the test and the weight provided to different historical time periods. The questions were not

intended to necessarily be representative of the difficulty level of the social studies subtest. A content analysis using Bulletin 2000 was not ideal, but it was the only option available. According to a personal email communication from Dr. Gloria Turner who served as the Director of Assessment for the Alabama State Department of Education, the content standards are “considered to be minimum, required, fundamental, and specific” (G. Turner, personal communication, February 11, 2008). Teachers familiar with the exam (through unsolicited student comments) also confirmed that the test is similar to the item specifications in the bulletin.

The content analysis was conducted using Bloom’s taxonomy and the first two authentic pedagogy standards associated with the task rubric (Construction of Knowledge and Elaborated Communication). The table at the end of this appendix provides a breakdown of how the 84 items were rated. All of the items were coded by three raters: Lamont Maddox, Dr. John Saye, and a graduate student trained with the AIW rubrics. At least two of the three raters agreed 94% of the time when applying Bloom’s taxonomy to categorize questions as having either low or high levels of cognitive difficulty. There was complete agreement among raters on the Elaborated Communication (EC) standard on the AIW rubric since the test is exclusively multiple-choice (and therefore does not require EC above a 1). Two of the three raters agreed 98% of the time when scoring based on the Construction of Knowledge standard on the AIW rubric. When disagreement occurred it was within one level on the rubric (1 or 2). In most instances the three raters agreed on the rating regardless of the method of analysis.

### *Results of Content Analysis*

Method of Analysis	N	%
Bloom's Taxonomy		
High	3	4%
Low	81	96%
Construction of Knowledge (AIW)		
Level 3	0	0
Level 2	11	13%
Level 1	73	87%

Note. Low = knowledge and comprehension; high = some application or analysis. Numbers reflect agreement among at least two of the three raters on each item.

#### Sample Items:

1. The first fort in America built by the Spanish was located in

- A. El Paso, Texas
- B. St. Augustine, Florida
- C. Natchez, Mississippi
- D. New Orleans, Louisiana

Scoring Notes: This task requires recall of factual information. It reflects a “low” knowledge level ranking on Bloom’s taxonomy since no higher order processes (i.e. synthesis, application, analysis, etc.) are required. On the construction of knowledge authentic intellectual work scale, this question most closely approximates a level 1 score whereby the dominate expectation is “that students will merely reproduce information gained by reading, listening, or observing.” The question also reflects a level 1 score for elaborated communication since it is multiple-choice.

2. The Missouri Compromise of 1820

- A. ended the slave trade in the United States.
- B. maintained a balance between slave and free states.
- C. granted political rights to slaves escaping to free states.
- D. allowed the expansion of slavery in all United States territories.

Scoring Notes: This question also requires recall of factual knowledge. In this case, students must remember what the Missouri Compromise accomplished. The question received a “low” knowledge level designation on Bloom’s taxonomy. On the construction of knowledge authentic intellectual work scale, this question most closely approximates a level 1 score whereby the dominate expectation is “that students will merely reproduce information gained by reading, listening, or

observing.” The question also reflects a level 1 score for elaborated communication since it is multiple-choice.

3. Use the passage below and your own knowledge to answer Number 5.

### **Removal of Southern Indians to Indian Territory, 1835**

The plan of removing the aboriginal people who yet remain within the settled portions of the United States...approaches its consummation...an extensive region...has been assigned for their permanent residence. It has been divided into districts and allotted among them. Many have already removed and others are preparing to go...

The pledge of the United States has been given by Congress that the [region] destined for the residence of this people shall be forever “secured and guaranteed to them.” A [region] ...has been assigned to them, into which the white settlements are not to be pushed...A barrier has thus been raised for their protection against the encroachment of our citizens...

The action described in the passage was a direct result of the

- A. growth of social reform movements.
- B. westward expansion of the United States.
- C. movement of people from rural to urban areas.
- D. acquisition of territories overseas by the United States.

Scoring Notes: This question also falls at the lower end of Bloom’s taxonomy, but it measures comprehension of the material in the paragraph instead of just recall. It scored at the “2” level on the construction of knowledge scale. This indicates that there “was some expectation for students to interpret, analyze, synthesize, or evaluate information, rather than merely to reproduce information.” The question also reflects a level 1 score for elaborated communication since it is multiple-choice.

## Appendix T: Notes on the Student Sample

Every attempt was made to make the student sample as inclusive as possible. Thirty students in the sample had multiple tenth grade social studies teachers because they took more than one social studies course during their tenth grade year. These students were excluded from the analysis for a number of reasons. In many cases, authentic pedagogy scores weren't available for both teachers. It was also difficult to isolate the effects of each teacher's instruction on student performance. With these students removed the resulting N for the study was 805. When factoring out students who did not have data listed for the social studies graduation exam variable, that number was reduced to 747.

The study schools have a strong reputation for academic excellence. As a result, they experienced a relatively high number of transfer students. Data was collected from the system at different points during the study as test results, student grades, and other information became available. The data collection schedule and changing student population at the schools produced some discrepancies in student documentation. For example, students were listed on class rosters (by ID #), but corresponding demographic or achievement data was not always available on the other spreadsheets. Whenever possible, I worked through the school system to resolve these differences. However, the final data set still had some missing data. The statistics for the various analyses included in chapter five are based on cases with no missing values for the variables used.



## Appendix U: Technical Description of Multiple Regression Analysis

Research Question Two: Do students that have been taught by teachers demonstrating higher levels of authentic pedagogy score higher on the Alabama High School Graduation Exam (AHSGE) than students taught by teachers with lower levels of authentic pedagogy?

The initial step in analyzing this question was to clearly define the predictor variables, other than authentic pedagogy, that most influenced students' graduation exam scores. I ran a series of regression analyses designed to filter out highly correlated variables that would overlap in explaining the variance of graduation exam scores. I first conducted a backward entry regression analysis in SPSS using demographic and course related variables. This procedure resulted in the removal of the course type predictor variable (courses that were a year or just a semester in length – Fall or Spring) because it was interacting with the course name variable (.895). A second regression was conducted sequentially using the forced entry method. Results of this analysis indicated that the course name variable (AP European History or Regular U.S. History) was highly correlated with authentic instruction (.802). This correlation was not very surprising since the sample of teachers was small, and the teacher with the highest authentic pedagogy score also taught the majority of the Advanced Placement courses. In order to better ascertain the impact of authentic instruction on the graduation exam results, the AP European History students needed to be removed from the analysis.

I ran a final sequential regression analysis and filtered out all of the students who took Advanced Placement European History. The resulting analysis included 427 students who took regular 10<sup>th</sup> grade United States history over the course of the two

years covered by the study. In evaluating the multiple regression models I made sure to test common assumptions. Ultimately, the assumptions needed for valid regression results were met.

## Appendix V: Technical Description of One-way ANOVA Procedures

In order to effectively run the ANOVA tests associated with research question two, the classes being compared need to be as similar as possible. The process that was used to pair like classes is described in this appendix. I first paired a class from the minimal authentic pedagogy category (Andy) with one from the limited authentic pedagogy category (Phillip). The classes included in the analysis were ones that I actually observed, although not in the same school year. Both classes were taught during the spring semester and were regular U.S. history courses. I compared the classes on specific demographic variables using the Pearson Chi-Square test. No statistically significant differences were found between the two classes in terms of gender, race, or socio-economic status. A subsequent T-test indicated that the limited authentic pedagogy class had higher mean social studies grades (87.55 vs. 82.50), but this difference was not significant ( $t=1.429$ ,  $p=.160$ ).

### *Comparison of Minimal and Limited Authentic Pedagogy Classes*

Variable	# of Students		Chi-Square Value
	Minimal	Limited	
Race			1.422
White	12	14	
African-American	11	6	
Gender			.023
Male	13	12	
Female	13	11	
SES <sup>a</sup>			1.243
Free/Reduced Lunch	3	6	
Paid	20	17	

*Note.* <sup>a</sup>Fisher's Exact Test (2-sided) consulted since two cells had expected count of less than 5. Result still did not reach significance (.459).

I also compared Andy’s minimal authentic pedagogy class with a class taught by the highest scoring tenth grade teacher (Ryan). Ryan’s class had higher mean social studies grades (85.45 vs. 80.92). The difference in means was not statistically significant ( $t=.924, p=.374$ ). The table below indicates that the classes did not differ significantly on gender or socio-economic status. However, they did differ based on race with the minimal class having significantly more African-Americans. As a result, I focused the subsequent ANOVA analysis on white students only.

*Comparison of Minimal and Moderate Authentic Pedagogy Classes*

Variable	Number of Students		Chi-Square
	Minimal	Moderate	
Race			6.571**
White	12	20	
African-American	11	3	
Gender			.087
Male	13	13	
Female	13	11	
SES			.505
Free/Reduced Lunch	3	5	
Paid	20	19	

*Note.* <sup>a</sup>Fisher’s Exact Test (2-sided) consulted since two cells had expected count of less than 5. Result still did not reach significance (.701).

\*\* $p < .01$ .

## Appendix W: Technical Description of Factorial MANOVA Procedures

This section provides additional information regarding the analysis of data associated with the higher order editorials (research question three). I will first discuss the process used to analyze the Manifest Destiny editorials before turning to the advanced placement German Unification writing task. In order to conduct the MANOVA analysis associated with this research question, the three groups (minimal, limited, and moderate authentic pedagogy) needed to be as similar as possible. I wanted to control for variables that could impact student performance other than authentic pedagogy. I compared the groups on specific demographic variables using the Pearson Chi-Square test. The results indicated that the groups did not differ significantly for gender or SES. However, the difference in race was significant. The three groups differed in their racial composition (black vs. white) beyond what you would anticipate happening by chance. The follow-up to this finding was to determine whether race played a significant role in influencing student performance on the higher order assessment. If it didn't, then the differences between the groups on this variable were irrelevant. I ran a MANOVA and found that race did have a significant impact on student performance (Hotelling's Trace  $p=.039$ ) so I decided to incorporate this variable into my final factorial MANOVA model which included the various authentic pedagogy teacher groupings (minimal, limited, moderate) as the other independent variables. This model revealed that there was not a statistically significant impact for race on student achievement on the designated dependent variables (Hotelling's Trace  $p=.107$ ).

In addition to trying to control for some of the demographic characteristics that could influence achievement on the Manifest Destiny higher order assessment, I also

conducted a one-way ANOVA to determine if the groups were significantly different in terms of students' grades in history. The assumption of homogeneity of variance was violated; therefore, the Welch F-ratio is reported. The teacher group (minimal, limited, or moderate) did not have a significant effect on student grades  $F(2, 70) = 2.047, p = .137$ . Put differently, the difference in mean 10<sup>th</sup> grade history average between the groups may be due to chance (82, 81, and 84).

*Comparison of Minimal, Limited, and Moderate Authentic Pedagogy Groups for Manifest Destiny Editorial*

Variable	Number of Students			Chi-Square
	Minimal	Limited	Moderate	
Race				9.320**
White	36	14	48	
African-American	23	13	11	
Gender <sup>a</sup>				.969
Male	35	14	30	
Female	27	15	32	
SES				5.461
Free/Reduced Lunch	13	10	9	
Paid	46	17	52	

*Note.* \*\* $p < .01$ .

A similar process was used to compare groups for the advanced placement editorial. In this case, there were only two groups: limited and moderate authentic pedagogy. The groups were similar in terms of gender, SES, and race. A T-test indicated that the limited authentic pedagogy class had lower mean social studies grades (83.40 vs. 84.29), but this difference was not significant ( $t = -.550, p = .584$ ). The results of these analyses suggested that a fair comparison could be made between the two authentic

pedagogy groups because no significant differences existed on the variables I had chosen to examine.

*Comparison of Minimal, Limited, and Moderate Authentic Pedagogy Groups for Advanced Placement German Unification Editorial*

Variable	Number of Students		Chi-Square
	Limited	Moderate	
Race <sup>a</sup>			
White	18	40	3.134
African-American	6	4	
Gender			
Male	10	27	2.742
Female	23	29	
SES <sup>b</sup>			
Free/Reduced Lunch	2	4	.048
Paid	31	51	

*Note.* <sup>a</sup>Fisher's Exact Test (2-sided) consulted since one cell had an expected count of less than 5. Result still did not reach significance (.148); <sup>b</sup>Fisher's Exact Test (2-sided) consulted since two cells had an expected count of less than 5. Result still did not reach significance (1.000).

## Appendix X: Higher Order Editorial Examples

### Advanced Placement Example #1

Part I	Scoring Notes
<p>Paragraph 1: Introduction.</p> <p>Germany has gone through uprisings/riots in 1848, won many wars including the Danish-Prussian and Austro-Prussian war; and has united northern Germany in 1867. German people should unify into one nation because the German people have shown their strength through winning many wars and proved that unification is possible when Northern Germany was united. Other nations and especially conservatives would oppose this because one German nation-state would be a threat to their country.</p>	<p>Position = 1. See underlined sentence in paragraph 2</p> <p>Context = 1. The student provides one sentence of historical background that sticks pretty close to the timeline.</p>
<p>Paragraph 2: Supporting Arguments</p> <p><u>The unification of all Germanic people should be endorsed by the German people</u> because we have proven to be a strong country united by winning many wars. The defeat of the Danish in 1864 shows that the German military is strong and can be very successful and victorious in battle creating a stable nation. To add to this, the unification of Northern Germany in 1867 shows that unification can be done. It would be in our best interest as German people to unite because the Constitution of the North German Confederation allows for more rights for citizens and there is hope for a constitution for all of Germany if we unify.</p>	<p>Persuasiveness= 3 “adequate.”</p> <p>The student never defines what a unified Germany would include (i.e. Austria?).</p> <p>Three main points are made. Unification will result in a stronger, more stable country. It can be accomplished (practical) and it would potentially bring more rights for citizens. The student doesn’t elaborate enough on these points for a four.</p>
<p>Paragraph 3: Opposing Views</p> <p>I think Conservatives in other countries would oppose the unification of Germany. Some arguments that any opposer would make could be that nationalism causes war in that countries/nations want to grow (Document 1). Along with this argument opposers might say that nationality is secondary as in document 1.</p>	<p>Dialectical Reasoning =0</p> <p>The student did not adequately describe the opposing view that nationalism causes war. Why would Germany “want to grow” and how would this lead to conflict? The other parts of this paragraph are also too brief to gauge student understanding.</p>
<p>Paragraph 4: Final Conclusion</p> <p>Although becoming unified may bring about wars to extend borders, with nationalism comes a sense</p>	<p>Quality of Final Position= 1</p> <p>The student responds to critics (even though the critic’s views</p>



---

of pride and the need to make your country strong from outside enemies. Take for example Northern Germany. Before 1867 they were scattered nations and when the Danish-Prussian War came along, they had no power to resist and were defeated. It is also a false accusation to say that nationality is secondary to other matters. Being one nation connect us all and make us feel obligated to make the country stronger. In conclusion Unification of all Germanic peoples is a great idea because it would make us part of a strong, stable nation, that is able to protect itself.

---

aren't very well described in paragraph 3) and offers a brief conclusion. The paragraph does not add to the persuasiveness of the editorial, especially since the example is not very clear and possibly inaccurate.

---

Part II

I think nation-states based on common ethnic groups, culture, or religion is a negative idea. Although it could bring relations among people of the same groups to grow stronger, it would support segregation and hatred among groups. If people can't interact with all sorts of diverse people, then there is no way of beginning to understand them. Not understanding people often brings the feeling of group superiority and anger towards others. I believe that experiencing different things and interacting with different people brings a more cultured and well rounded society that promotes peace and understanding not segregation and hate.

---

Scoring Notes

Decision-Making = 0 – Never mentions anything about U.S. policy although the student's view can be inferred.

Persuasiveness = 2 – "Adequate" The response has a chance of persuading the reader. The student argues that new nation-states would promote more hatred and increase the likelihood that groups will not understand each other. The ideas are not supported with any elaboration.

---

Advanced Placement Example #2

Part I	Scoring Notes
<p>Paragraph 1: Introduction.</p> <p>The unification of Germany would only escort a plethora of problems and end up dumping upon us strife. Germany has been embedded in a perpetual state of war. [We've endured the Danish-Prussian War, Austro-Prussian War, and the Franco-Prussian War, and we German people are drained like our economy]. It will be impossible to unite peacefully and successfully the currently fragmented Germany. <u>German[s] should oppose the unification of Germany</u> because of our economic stature of Germany as well as the impossibility of uniting all these prideful German nationalities.</p>	<p>Position = 1 – The student clearly is opposed to German unification.</p> <p>Context = 1 – Some historical context provided in the bracketed sentence.</p>
<p>Paragraph 2: Supporting Arguments</p> <p>If Germany unifies, we will never exit a state of war. Germany is composed of numerous nationalities who own a sense of entitlement. No nationality will want to engage the compromise that will be required by unification (1). Not only that, but we would have to annex parts of Switzerland, the Netherlands, and Belgium. These countries certainly will not be happy and will most likely pursue war (2). Another concern plaguing my mind is our economic situation. Germany has been engaged in war after war, and we all know war drains economies (3). Trying to unify in a time of economic difficulty is certainly not a smart move. Because we are weary of war and our economy is drained, it is not a good idea that Germany should unify.</p>	<p>Persuasiveness =3 “Adequate”</p> <p>The student provided three arguments with one of them closely following the source document. None of the arguments include much in the way of support. The economic argument may be correct, but the student provides no information to substantiate the claim. The editorial would be more persuasive if the student acknowledged the possibility of more limited forms of unification.</p>
<p>Paragraph 3: Opposing Views</p> <p>Faulty arguments are flung at my opposition to German unification. <u>My critics claim that all Germany has a sense of unity after fighting together to defeat France, and thus should unify.</u> Although this is true, everyone must keep in mind that this sense of unity was during a time of war where we were all seeking to defeat France. This sense of unity will soon fade and be replaced by arguing nationalities each fighting for their own good. The conflict</p>	<p>Dialectical Reasoning =2</p> <p>The student provided two feasible opposing views. Both of them are not described in much detail. The student also didn't address some significant points that would hurt</p>

with France is done and soon Germany will not have a common interest to fight for. Other critics say that together, Germany could rise and be a great power who eventually controls the world. This is an absolutely absurd notion. Because of the aforementioned nationality conflict, Germany will have too much internal conflict to focus on international affairs. And do not forget the troublesome question of whether Austria would be able to join a unified Germany. My critics arguments are superficial and need serious reevaluation.

his/her argument. For example, the possibility that unification could *stimulate* economic growth.

Paragraph 4: Final Conclusion

Obviously, German unification is a bad idea. If Germany attempts to unify all these different nationalities during our dismal economic condition, we will never exit a state of war. The only reason we Germans should unify is against the very idea of German unification.

Quality of Final Position= 1  
The student addressed the views of critics in the preceding paragraph. He/she offers a short conclusion that restates the economic argument. It does not add (or detract) from the persuasiveness of the editorial.

---

Part II

The U.S. is faced with a difficult policy decision when it comes to the support of nation-states based on cultural identity. In the real world, everyone cannot always be happy. There must be compromise. Nobody can always have their way. I believe that the U.S. should respect each and every culture's rights and safety and if these basic rights and safety can only be obtained by creating a new nation-state, then I think the U.S. should support it. But, a far better route for the U.S. to take is supporting compromise and peace in an existing nation. Every nation is going to encounter problems, even the nations that break off because of basic rights and safety. Simply amputating a cultural group from a mother nation is not going to solve all the problems. So, if the U.S. supports peace and compromise in existing nations, people will learn to live peacefully with other

1,1 No concrete examples. All countries experience problems- people will learn to live peacefully with each other.

---

people as opposed to speedily and selfishly forming a cultural bubble. Altogether, I think the U.S. should support the unity of a nation as opposed to many separate nation-states.

---

U.S. History Higher Order Editorial Example 1

Part I	Scoring Notes
<p>Paragraph 1: Introduction.</p> <p>The war with Mexico has been going on for a year now, and many people still have differing opinions on the matter. Texas was full of settlers from our country, and the Mexican government tried forcing many unfair laws upon them. They fought for their independence and won it, but the Mexican government refused to acknowledge their clear victory. When the new, independent Texas tried to join us, the Mexican government ignores that and us. I think we are perfectly justified in going to war with Mexico for many reasons.</p>	<p>Position = 1 – The student clearly supports Manifest Destiny based on the last sentence of the editorial.</p> <p>Context = 2 – The student accurately describes some of the events leading to the decision-point and does so without simply copying the timeline.</p>
<p>Paragraph 2: Supporting Arguments</p> <p>If we win the war with Mexico, the people we liberate from their controlling government will only benefit. We didn't decide to start violently pursuing our Manifest Destiny either. Mexico forced us into it by not acknowledging Texas independence and freedom to choose who they wish to follow. The Mexican government is too stubborn to see that everyone could benefit from us pursuing our Manifest Destiny. We were trying to buy California and New Mexico from them.</p>	<p>Persuasiveness = 3 "Adequate". The editorial has a chance of persuading the reader. The student provides at least two reasons why U.S. actions were justified. However, the reasons have little to do with Manifest Destiny &amp; they are not supported with enough evidence to warrant a higher score.</p>
<p>Paragraph 3: Opposing Views</p> <p>Some say that the only reason we are at war with Mexico is because <u>we are greedy</u>. Others think we are only at war because we <u>didn't consider the rights of other countries</u>. We would be gaining much land, and at the expense of Mexico, if we win this war.</p>	<p>Dialectical Reasoning = 2 The student dedicated a paragraph to opposing views without responding (which was rare). Two points are made in minimal detail.</p>
<p>Paragraph 4: Final Conclusion</p> <p>To those that think to United States is being greedy, consider the facts. We were trying to buy territory from Mexico before this war started. Texas no longer belongs to Mexico, so we weren't violating any of their rights, and Texas wanted to become a part of our country. Pursuing Manifest Destiny is something we should do, but only if it is done without violating any rights. Because Texas is not a part of Mexico, because Mexico is overly controlling, and because it is our destiny, war with Mexico is</p>	<p>Quality of Final Position= 2 "Approaching Satisfactory" The student did a decent job of reiterating points made earlier in the editorial. The paragraph as a whole added to the persuasiveness of the editorial, but it did not represent advanced</p>

completely justified.

dialectical reasoning. The student didn't appear to thoughtfully consider the validity of the opposing views introduced in paragraph three. They were basically dismissed out of hand.

---

---

## Part II

---

I think the U.S.'s mission is to be the "police." We tend to take care of people that are having their rights taken away. I'm not sure how this could be accomplished for the whole world, but I think that some day, a long time from now, it could be accomplished.

---

U.S. History Higher Order Editorial Example 2

Part I	Scoring Notes
<p>Paragraph 1: Introduction.</p> <p>In my opinion using the “saying” Manifest Destiny does not justify the war. In the case that it does, that means that anything wrong someone can just say that some how they know that God wanted them to do it. We took the wrong approach in getting Texas. We should have paid for it if we wanted it and not killed hundreds of people.</p> <p>The situation between the United States and Mexico was that the United States wanted Mexico and Santa Anna did not want to give it up. Some of the important events were the four battles that finally won over Texas to the United States. <u>The United States is wrong to use Manifest Destiny to go to war with Mexico.</u></p>	<p>Position = 1 – Clearly stated at the end of the introduction.</p> <p>Context = 1 – The student provides some context, but it is not at all clear whether he/she really understands the events that preceded the Mexican-American War. This was borderline “0”.</p>
<p>Paragraph 2: Supporting Arguments</p> <p>Mexican problems are Mexicos problems. The United States should first worry about the wrong going on in the United States. Just because we want something doesn’t mean that we can just take it.</p>	<p>Persuasiveness = 2</p> <p>“Minimal” – The editorial is not likely to persuade the reader. The main points can essentially be summarized as might doesn’t make right and “we” (the U.S.) can’t always have what we want. The explanation is not clear enough to warrant a higher score.</p>
<p>Paragraph 3: Opposing Views</p> <p>Some people might disagree with me because they think that something is going wrong over there, but there are things that are going wrong in the United States. Slavery is one of the biggest issues. If you think about it we are doing the same thing as them. Also that means that any country in the world can just come over and decide that they want a part of our country and if they are better fighters then us then they just win our country? Its not right.</p>	<p>Dialectical Reasoning = 0</p> <p>The student doesn’t accurately provide an opposing view. However, this was one of the few editorials to mention slavery.</p>
<p>Paragraph 4: Final Conclusion</p> <p>So there are my decisions to what I think about the Manifest Destiny and why I think what I think.</p>	<p>Quality of Final Position= 0</p> <p>The student doesn’t restate any arguments.</p>

---

Part II	Scoring Notes
<p>My opinion on the war is that what those people over there do is their business. America already has so many problems without getting in everyone elses. There are millions of kids without homes or food. Millions of kids dropping out of school, loss in jobs, people killing people, but we are too busy in everyone elses problems. America is a free country and I think that was our mission and destiny.</p>	<p>0 = The student gets sidetracked in focusing on the war (Iraq?). The last sentence ties back into the question some, but I really couldn't determine if the student saw any connections between America's historic sense of Manifest Destiny and its place in the world today. The overall isolationist, America first, stance was a common one expressed by students in this part of the assessment.</p>

---