Authentic Pedagogy and the Acquisition of Lower Order Knowledge in History

Lamont E. Maddox  
*University of North Alabama*

John W. Saye  
*Auburn University*

This study examined the impact of varying levels of authentic pedagogy on student learning in select 9th and 10th grade history classrooms. The sample included four junior high 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 that was used as the primary independent variable in subsequent statistical analyses designed to evaluate student learning outcomes. The findings suggest that the use of authentic tasks and instruction has a small, but positive correlation with student performance on the Alabama High School Graduation Exam in use at the time the study was done. A performance benefit was also noted for students who experienced multiple courses at the moderate authentic pedagogy level. The benefit, however, could be attributed to an advanced placement effect since advanced placement students in the sample were more likely to receive moderate authentic pedagogy.

Key Words: Authentic Intellectual Work, Standardized Testing, Inquiry, Assessment, Secondary Social Studies, U.S. History

Introduction

Social studies teachers in today’s high-stakes testing environment often face conflicting demands. Graduation exams and end of course tests tend to measure the extent to which students master a specific body of factual information (Grant, 2006; DeWitt et al., 2013). At the same time, educators are expected to provide lessons that increase the capacity of students to engage in the critical thinking required for 21st century citizenship (Alabama Department of Education, 2011; Partnership for 21st Century Skills, 2011). To help students attain higher order outcomes, researchers often advocate inquiry-based instruction (Saye & Brush, 2004). A variety of impediments, however, to this type of instruction exist in public school settings (Onosko, 1991). In trying to reconcile the need to prepare students for high-stakes tests while also pursuing higher order objectives, teachers are confronted with a persistent dilemma in curricular decision making: depth vs. breadth (Newmann, 1988; Salinas, 2006; Sizer, 1984; van Hover, 2006).

Many teachers are hesitant to sacrifice content coverage in order to provide more time intensive inquiry lessons.

History teachers in Alabama encountered this dilemma when preparing students for the U.S. History section of the Alabama High School Graduation Exam (AHSGE), a test that measured mastery of standards “considered to be minimum, required, fundamental, and specific” (G. Turner, personal communication, February 11, 2008). The AHSGE was phased out recently in favor of an end of course test in U.S. History, but budget constraints have prevented
implementation of this new assessment. This study is still of value since many teachers encounter the depth versus breadth dilemma in states that continue to use accountability systems based on high-stakes tests that are very similar to the AHSGE in terms of their emphasis on measuring student mastery of lower-order knowledge (see DeWitt et al., 2013). Even some AP teachers experience this problem if their course features an exam with a substantial multiple-choice component and heavy emphasis on acquisition of specific content (Parker et al., 2011). Minimum competency tests like the AHSGE exert pressure on teachers to “teach to the test” by encouraging them to narrow the curriculum (Au, 2007) and rely on traditional forms of assessment. Many administrators and teachers are unlikely to invest the time and resources needed to increase inquiry-based instruction in history classrooms, without some assurance student performance will not suffer on those high stakes tests (Grant et al., 2002).

In the present study, we sought to determine the relationship between inquiry-based instruction (authentic pedagogy) and student performance on the Alabama High School Graduation Exam. The level of inquiry students experienced in their classrooms was operationalized using Fred Newmann, Bruce King, and Dana Carmichael’s (2007) authentic intellectual work (AIW) framework. Teachers who use inquiry-based instruction in a manner consistent with this framework challenge their students to “construct knowledge, through disciplined inquiry, to produce discourse, products, or a performance that has value beyond school” (Newmann, King, & Carmichael, 2007, p. 3).

**Literature Review**

Advocates of inquiry-based instruction argue that active, inquiry-based activities do not preclude the acquisition of lower order content knowledge (Gallagher & Stepfen, 1996; Newmann, Bryk, & Nagaoka, 2001; Saye & Brush, 1999). In science and mathematics, there is evidence to support this notion at virtually all levels of schooling (AIR/SRI, 2006, 2007; D'Agostino, 1996; Lee, Smith, & Croninger, 1997). Research in social studies, however, is less definitive despite some noteworthy recent studies (Parker et al., 2011; Saye & SSIRC, 2013; Smith & Niemi, 2001; Wirkala & Kuhn, 2011). Social studies research on the impact of inquiry-based instruction has been plagued by the use of inconsistent terminology, poor research design, and a variety of other factors (Hahn, 1991; Metcalf, 1963; Newmann, 1991; Onosko, 1991). The use of the AIW framework may mitigate some of these problems. It enables researchers to more precisely describe the intellectual challenges experienced by students in different classroom environments. As a result, consumers of the research can more readily evaluate the merits of the research findings comparing various instructional approaches.

Most of the studies using the AIW framework have investigated the extent to which authentic pedagogy elicits authentic, higher order outcomes. This line of research, involving multiple disciplines and diverse student populations, has consistently shown a strong relationship between high levels of authentic pedagogy and the production of high quality student work (Amosa, Ladwig, Griffiths, & Gore, 2007; Avery, 1999; Avery, Freeman, Carmichael-Tanaka, 2002; King, Schroeder, and Chawswczewski, 2001; Koh, Kim & Luke, 2009; Ladwig, Smith, Gore, Amos, & Griffiths, 2007; Newmann & Associates, 1996; Newmann, Lopez, & Bryk, 1998). The studies that examined the impact of authentic pedagogy on the acquisition of lower order knowledge are much more limited and focused on subjects other than social studies.
Perhaps the most significant study to analyze the relationship between AIW and student performance on standardized tests was conducted in Chicago with a sample of 19 schools (Newmann, Bryk, & Nagaoka, 2001). This study analyzed reading, writing, and math performance using the Iowa Test of Basic Skills (ITBS) and the Illinois Goal Assessment Program (IGAP). Researchers found that eighth grade students in classes that received higher scoring authentic tasks outperformed their peers on the writing and math portions of the ITBS and on the math, writing, and reading sections of the IGAP (Newmann, Bryk, & Nagaoka, 2001).

A study by Valerie Lee, Julia Smith, and Robert Croninger (1997) investigated a similar question as part of an effort to determine the factors that most contribute to the effectiveness of small communal schools. Analyzing data from the National Educational Longitudinal Study, they found that students in schools with higher levels of authentic instruction made larger gains on the science and math achievement tests (Lee, Smith, & Croninger, 1997).

A more recent study, using modified AIW rubrics, yielded less definitive results. The High School Grants Initiative included two studies that analyzed the extent to which assignments featuring high levels of rigor and relevance influenced the quality of work produced by students in English Language Arts (ELA) and Math. Researchers noted a significant positive relationship between quality student work (as defined by AIW principles) in ELA and improved standardized test scores in reading. The relationship in Math, while positive, did not reach statistical significance (AIR/SRI, 2006). In a later study, researchers found a statistically positive relationship between quality student work and math scores, but there was not a positive relationship for language arts (AIR/SRI, 2007). The AIW model has also influenced education reform initiatives in Australia. In a study of schools that instituted a New Basics curriculum emphasizing higher order authentic instruction and the use of authentic tasks to assess student learning, investigators found no decline in standardized achievement test scores in literacy and numeracy (Education Queensland, 2004).

The evidence presented thus far suggests a mostly positive relationship between AIW and student performance on lower order knowledge high-stakes tests in subjects other than social studies. In social studies, two AIW studies have included a focus on lower-order outcomes. A study conducted by Patricia Avery (1999) with a small sample of U.S. history teachers investigated how varying levels of authentic instruction impacted students’ ability to complete a common higher-order authentic task. The study included a lower-order multiple-choice test as a secondary achievement measure. The results of Avery’s study yielded a small, positive statistical link between authentic pedagogy and performance on this achievement test (Avery, 1999). The Social Studies Inquiry Research Collaborative (SSIRC), more recently, completed a larger six-state study of 93 classes to determine whether students who experienced higher levels of authentic pedagogy in their social studies classes performed better on high-stakes tests. The SSIRC study found that higher levels of authentic pedagogy were associated with higher scores on state mandated tests (Saye & SSIRC, 2013). A final study corroborates these findings, although it did not explicitly utilize the AIW rubrics. In this study, Walter Parker and his colleagues (2011) conducted a design experiment with a project-based AP Government and Politics course that engaged students in authentic intellectual work. The researchers conducted two studies to determine if the redesigned course would allow students to perform as well or better than comparable students in traditional AP classes on the end of course AP exam (Parker et al., 2011). The students who experienced the project-based curriculum “scored as well (study
2) or significantly better (study 1) on the AP exam than students in Traditional-AP classes” (Parker et al., 2011, p. 553).

An important question that has received no attention in prior AIW research is whether performance on high-stakes tests improves when students are exposed to multiple courses that feature higher levels of authentic pedagogy. Although no AIW research or general social studies research has examined this question, some relevant work has been done in science. A study conducted by Michael Klentschy, Leslie Garrison, and Olga Maia Amaral (2001) 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.

Our research continues the AIW line of inquiry by seeking to establish the applicability of the findings from the Fred Newmann, Anthony Bryk, and Jenny Nagaoka (2001) study to history classrooms while also asking whether achievement gains increase when students experience multiple classes at the moderate authentic pedagogy level.

**Conceptual Framework**

Authentic pedagogy focuses on instructional practices designed to elicit authentic intellectual work from students. The AIW model has been refined and field-tested over several decades (Newmann & Archbald, 1988; Newmann, Bryk, & Nagaoka, 2001; Newmann, King, & Carmichael, 2007; Newmann, Secada, & Wehlage, 1995; Saye & SSIRC, 2013). The standards associated with this model were devised after considering the types of intellectual challenges commonly faced by adults in their daily occupations. 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). Construction of knowledge requires students to move from being consumers of information to producers who use prior knowledge to construct new interpretations or solutions to problems. To ensure solutions have rigor, students engage in disciplined inquiry, using procedures and rules of evidence that are considered legitimate by professionals in the academic discipline under study. A disciplined approach to inquiry also requires students to convey their findings to others through elaborated forms of communication that provide deep, nuanced explanations of their work. Finally, authentic intellectual work has value beyond school because it 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 may include students attempting to influence public policy by writing and sending a persuasive letter to a Congressman.

**Design**

The purpose of this study was to examine the relationship between authentic pedagogy and student performance on the Alabama High School Graduation Exam. Our study was the pilot for the larger SSIRC project (Saye & SSIRC, 2013) and data collected from the Alabama sites in this study were incorporated into the results. The present study is important because it
permits comparisons to be made between what happened at our sites in Alabama and the overall sample of students from several states. Our study asked:

1. Do students who 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?

2. Does performance on the Alabama High School Graduation Exam (AHSGE) improve with repeated exposure (multiple courses) to classroom experiences that require students to perform challenging intellectual tasks?

The study utilized a mixed-methods, correlational research design to investigate existing instruction at two study schools over the course of two school years (2007-2008 and 2008-2009). We first assessed the level of authentic pedagogy students experienced in the history classrooms that were included in the study by applying AIW rubrics to data collected during classroom observations (field notes & assessment tasks). We then used statistical measures to analyze the relationship between higher teacher scores on the authentic pedagogy continuum and student achievement gains on the AHSGE.

**Setting and Participants**

The study took place in a junior high school and high school in the same school system in southeastern Alabama. The high school’s enrollment was 1,156 students while the junior high school had 908 students. Both schools were similar in terms of the diversity of their student bodies. The high school was 61.6% White, 30.6% Black, 4.8% Asian/Pacific Islander, and 1.8% Hispanic. Student enrollment at the junior high school was 60.3% White, 30.5% Black, 6.4% Asian/Pacific Islander, and 1.8% Hispanic (CCD Public School Data, 2007-2008).

Approximately 20% of the students received a free or reduced lunch. The selection of these schools represented purposeful sampling. Previous work in these settings suggested that some teachers in each school might be expected to score relatively highly on the AIW rubrics, thereby resulting in a more diverse range of scores on the authentic pedagogy continuum. The entire 9th and 10th grade social studies faculty volunteered to serve as participants \(N=8\). All teachers were White and most were male (6 of the 8). The majority taught for at least ten years (6 of the 8) and held advanced degrees (5 of the 8). The four junior high school teachers taught 9th grade World History exclusively. The remaining 10th grade high school teachers primarily taught U.S. History although two also taught the AP European History course.

**Instruments**

Rubrics developed by Newmann and his associates measure the three subcomponents of authentic pedagogy: instruction, tasks, and student work (Newmann, Secada, & Wehlage, 2009). The AIW instruction and task rubrics were used in the present study. The authentic task rubric enabled us to analyze the extent to which assigned tasks featured construction of knowledge, elaborated communication, and a connection to students’ lives. The instruction rubric measured the extent to which students experienced instruction that elicited higher order thinking, depth of knowledge, substantive communication, and a connection to life outside the classroom. Scores on the task and instruction rubrics were combined to assign an overall authentic pedagogy score for each teacher (possible range of 7-30). We conceptualized authentic pedagogy as a continuum. Teachers on the lower end of the continuum used a great deal of didactic instruction. As the scores increased, they represented greater use of authentic pedagogy (in-depth analysis of
In order to increase reliability, we received training with other investigators from the broader SSIRC study (Saye & SSIRC, 2013) on the use of the AIW rubrics by Bruce King, a member of Newmann’s research team. The SSIRC steering committee (of which we were a part) later field tested the rubrics and modified them slightly to clarify scoring for social studies classrooms. Several norming sessions were conducted among all the SSIRC researchers to ensure that the rubrics were being applied in a consistent fashion.

**High-stakes test.** Student achievement was measured using results from the Alabama High School Graduation Exam. The exam was administered for the first time in the tenth grade and included content from the beginning of U.S. History through World War II. The main concern in using this test for the study was the possibility of it measuring only lower order thinking and knowledge. The Alabama State Department of Education did not release test questions or prior versions of the test. A content analysis was conducted using an item specifications bulletin provided by the state containing 84 sample multiple-choice questions (Richardson, 2000). The analysis confirmed that the vast majority of the questions measured basic knowledge or comprehension of information (Patterson et al., 2012).

**Data Collection and Analysis**

We collected both teacher and student data. Teacher data consisted of instructional tasks, classroom observations, and interviews. Each teacher initially was asked to submit three challenging tasks that provided the best evidence of students performing their subject at the highest levels. Observations then were conducted to coincide with the period when students would be engaged in work related to each of the tasks. A brief interview was conducted with each teacher to better understand the intent behind each task. The interview also permitted us to develop a demographic profile of the study participants. Task and observation data were scored using the AIW rubrics. The reliability of AIW ratings was assessed by subjecting 25% of the tasks and 22% of the classroom observations to review by a second researcher. We obtained exact agreement on the various components of the instruction rubric 65% of the time. Exact agreement on the task rubric was higher at 83%. Inter-rater reliability was 100% when the standard was exact agreement or off by 1. These ratings compare favorably with reliability standards in prior AIW research of greater than 65% exact agreement and greater than 90% agreement within 1 point (Newmann & Associates, 1996).

In the second phase of the study we collected student data consisting of prior social studies grades, graduation exam results, and demographic data. This information was coded by the school system and sent to the researcher in a format that maintained the anonymity of the students. This procedure enabled us to include in the research database all tenth grade students who took social studies courses during the 2007-2008 and 2008-2009 school years.

The first step in data analysis was to quantify the extent to which teachers provided authentic pedagogy. The AIW rubric scores for the three tasks were averaged to develop an overall authentic task score. The three associated instruction scores also were averaged to create an authentic instruction value. The resulting task and instruction scores were combined to create an overall authentic pedagogy score that could range from a minimum of 7 points to a maximum of 30 points. Cut scores were used, based on quartiles created by the SSIRC (Saye & SSIRC, 2013), to form four descriptive categories representing different levels of authentic pedagogy. Teachers in the sample were described as utilizing minimal authentic pedagogy if they scored under 12, limited if they scored between 12 and 17.99, and moderate if they scored between 18
and 23.99. The final category (scores of 24 and above) represented substantial use of authentic pedagogy. In general, as scores increase along the continuum from minimal to substantial authentic pedagogy, instruction should gradually shift from an almost exclusive focus on the transmission of lower order knowledge to a curriculum that provides multiple opportunities for students to apply their disciplinary knowledge in analyzing real world problems. These categories were used as the basis for an analysis of the impact of various levels of authentic pedagogy on select student performance on the Alabama High School Graduation Exam (AHSGE).

Student achievement data from the Alabama High School Graduation Exam (AHSGE) was subjected to statistical analysis. AHSGE test results were organized into raw scores and scale scores. The raw score reflected the actual score students achieved on the 100 question test. The results, once scaled, range from 200 to 800 points (S. Dubose, personal communication, 2006). Scale scores were used to measure student achievement.

Two statistical approaches were used to investigate the impact of authentic pedagogy on graduation exam performance: multiple regression with students as the unit of analysis and ANOVA tests comparing specific classes. In conducting the 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, the final model included 427 students who took regular 10th grade United States history over the course of the two years covered by the study.

The class level analysis (ANOVA) required selecting two similar classes for comparison from teachers who utilized different levels of authentic pedagogy. Statistical tests (Pearson Chi-square) were performed to ensure the two classes did not differ significantly on key variables likely to influence achievement (i.e. demographics, prior achievement, etc.). In comparing the minimal authentic pedagogy class with the limited class, no significant differences were found in terms of gender, race, SES, or prior social studies grades. When the minimal class was compared with the moderate class, however, the difference for race (White vs. Black) reached statistical significance ($p < .01$). To avoid capturing differences in achievement that were mainly due to the confounding variable of race, the final ANOVA only included White students from each class. All other variables (gender, SES, prior achievement) were similar for students in these two classes.

The second research question focused on whether repeated exposure (multiple classes) to moderate levels of authentic pedagogy was associated with higher scores on the Alabama High School Graduation Exam. This aspect of the study was addressed by including both ninth and tenth grade history teachers in the study thus enabling us to analyze the instruction students received during the year when the graduation exam was first administered (grade 10) as well as the instruction they experienced in the preceding year. We collected teacher data for the entire sample of eight teachers during the first two semesters of the study. Although we were unable to observe firsthand the instruction participating students received in their 9th grade history classes (the year before the study began), we made the assumption teachers are unlikely to radically alter their instruction between semesters or consecutive school years. The work of Larry Cuban (1984) and Dan Lortie (1975) supports the durability of established teachers’ practice. Based on this assumption, we used 9th grade teacher AIW scores gathered during the two semesters of the study as indicators of the level of authentic pedagogy that students experienced from those same
9th grade teachers in the previous year. We matched these AIW scores with the individual student records of each student who took the AHSGE in 2008 and 2009.

The second research question was investigated using a one-way ANOVA. A new prior moderate variable was created representing the total number of social studies courses each student had at the moderate authentic pedagogy level. Each student record in the research database already included a field containing the authentic pedagogy score for their ninth and tenth grade social studies teachers. These fields were used to calculate the prior moderate values for each student. Over the course of the 9th and 10th grade, a student could have 0, 1, or 2 social studies classes at the moderate level. Students who had more than one social studies teacher in any particular year were filtered out of the analysis. With the data organized in this fashion, we were able to use the prior moderate variable to investigate the relationship between multiple courses at the moderate authentic pedagogy level and student achievement on the AHSGE.

Results

This study investigated the relationship between the level of authentic pedagogy experienced by students in their history classes and their performance on the Alabama High School Graduation Exam. We first present an overview of the level of authentic pedagogy exhibited by the eight study teachers. We then examine the relationship between authentic pedagogy and student achievement. We conclude this section by examining the “compounding effect” of moderate levels of authentic pedagogy on student achievement.

Instruction in Study Classrooms

High levels of authentic pedagogy were rare in the teacher sample (see Table 1); this finding is consistent with similar studies (Ladwig, Smith, Gore, Amosa, & Griffiths, 2007; Newmann & Associates, 1996; Saye & SSIRC, 2013). The sample of teachers in the SSIRC study, by way of comparison, featured 21% at the moderate level and no teachers at the substantial authentic pedagogy level (Saye & SSIRC, 2013). The SSIRC researchers maintain moderate levels of authentic pedagogy reflect teaching that most educators would recognize as strong and challenging instruction. In the present study, there was enough variation across levels to determine whether higher levels of authentic pedagogy are associated with higher levels of student achievement.

Table 1

<table>
<thead>
<tr>
<th>Quartile</th>
<th>AIW Scores</th>
<th># of Teachers</th>
<th>% of Teachers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimal</td>
<td>7-11.99</td>
<td>3</td>
<td>37.5</td>
</tr>
<tr>
<td>Limited</td>
<td>12-17.99</td>
<td>2</td>
<td>25</td>
</tr>
<tr>
<td>Moderate</td>
<td>18-23.99</td>
<td>3</td>
<td>37.5</td>
</tr>
<tr>
<td>Substantial</td>
<td>Above 24</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Note. N=8.

Authentic Pedagogy and Graduation Exam Performance

The first research question asked: Do students who 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? This question was first examined using a regression analysis based on a sample of students ($N = 427$) who took U.S. History in the tenth grade during the two years of the study. Several predictor variables were incorporated into the analysis (see Table 2) resulting in an overall model with a large effect ($R^2 = .437^{***}$, $p = .001$). The demographic variables contributed the most in explaining the variance in scores (26%) followed by students' prior achievement, based on their tenth grade social studies average (15%). In both instances, the contribution was statistically significant. When the authentic pedagogy variables were added to the model they provided a small, statistically significant positive effect of 3% in promoting student achievement. Inspection of the squared semi-partial correlations for authentic instruction and authentic tasks yielded contradictory results. The unique contribution of authentic instruction had a positive effect on student graduation exam scores ($r_{sp}^2 = .021$, $p < .001$). However, the influence of authentic tasks was negative ($r_{sp}^2 = .018$, $p < .001$), but not very strong compared to other predictor variables. Some possible explanations for this finding are included in the discussion section.

Table 2

<table>
<thead>
<tr>
<th>Model</th>
<th>Variable</th>
<th>$R^2$</th>
<th>$\beta$</th>
<th>Semi-partial$^c$ (percentage)$^d$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Demographics</td>
<td>Gender</td>
<td>.261***</td>
<td>.235***</td>
<td>.229*** (5.2%)</td>
</tr>
<tr>
<td></td>
<td>Ethnicity</td>
<td>.189***</td>
<td>.139***</td>
<td>(1.9%)</td>
</tr>
<tr>
<td></td>
<td>LEP</td>
<td>.047</td>
<td>.047</td>
<td>.22%</td>
</tr>
<tr>
<td></td>
<td>Special Ed</td>
<td>-.141***</td>
<td>-.138***</td>
<td>(1.9%)</td>
</tr>
<tr>
<td></td>
<td>SES</td>
<td>.068</td>
<td>.050</td>
<td>(2.5%)</td>
</tr>
<tr>
<td>2. Achievement</td>
<td>10th Average</td>
<td>.149***</td>
<td>.444***</td>
<td>.380*** (14.4%)</td>
</tr>
<tr>
<td>3. Authentic Pedagogy</td>
<td>Task</td>
<td>.027***</td>
<td>-.151***</td>
<td>-.134*** (1.8%)</td>
</tr>
<tr>
<td></td>
<td>Instruction</td>
<td>.163***</td>
<td>.144***</td>
<td>(2.1%)</td>
</tr>
<tr>
<td>OVERALL MODEL$^b$</td>
<td></td>
<td>.437***</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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

$^a$The regression model was statistically significant (F = 40.48, $p < .001$). Using effect size guidelines suggested by Cohen (1988) and Pedhazur (1997), the $R^2$ of .437 can be classified as a large effect.

$^b$Beta ($\beta$) is the standardized weight used to represent the strength of the predictor variable's contribution toward the prediction of the outcome (achievement).

$^c$Semi-partial correlation is used to represent the unique relationship between the predictor and the outcome.

$^d$This is the squared semi-partial correlation. This represents the unique contribution of the predictor variable, or the proportion of the total variation in the outcome (achievement) that is accounted for by the individual predictor variable.
In addition to the regression analysis, we used a one-way ANOVA design to compare the performance of students in intact classes featuring various levels of authentic pedagogy. This approach enabled us to more precisely determine if students in classes with higher authentic pedagogy outperformed their peers in similar classes with lower authentic pedagogy. When pairing a minimal authentic pedagogy class with a class from the limited authentic pedagogy category, the minimal authentic pedagogy class performed significantly better, $F(1, 44) = 9.516$, $MSE = 2591.5$, $p = .004$, $\eta^2 = 0.18$. This finding should be considered with some caution since the two teachers were fairly close together on the authentic pedagogy scale (10.9 compared to 13.3). Also, it is possible that uncontrolled variables played a role in contributing to the difference in outcomes (i.e. variables such as teaching experience, etc.).

A minimal authentic pedagogy class also was compared with a class taught by the highest scoring tenth grade teacher in the moderate category. Both classes were regular U.S. History courses with similar students except for their racial composition. We focused our analysis for the minimal/moderate comparison on White students only in order to control for this difference. The analysis revealed no difference in graduation exam performance, $F(1, 29) = .000$, $MSE = 3033.055$, $p = .986$. Since the cut score on the graduation exam was a 509 and the mean score of the moderate class was 558, White students in the moderate authentic pedagogy class seemed to perform relatively well on the AHSGE which was a test of lower-order knowledge (see Table 3).

### Table 3

**Comparing Graduation Exam Performance for Minimal, Limited, and Moderate Authentic Pedagogy Classes**

<table>
<thead>
<tr>
<th>Class</th>
<th>Mean SS AHSGE</th>
<th>SD</th>
<th>Range</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Limited AP</td>
<td>512.27</td>
<td>46.809</td>
<td>491.52 to 533.03</td>
<td>$F=9.516$</td>
</tr>
<tr>
<td>Minimal AP</td>
<td>558.62</td>
<td>54.380</td>
<td>535.66 to 581.59</td>
<td></td>
</tr>
<tr>
<td>Moderate AP</td>
<td>558.45</td>
<td>51.855</td>
<td>534.18 to 582.72</td>
<td>$F=.000$</td>
</tr>
<tr>
<td>Minimal AP</td>
<td>558.82</td>
<td>60.720</td>
<td>518.03 to 599.61</td>
<td></td>
</tr>
</tbody>
</table>

*Note. This table reflects two separate ANOVAs. The first compared limited and minimal AP classes. The second compared a moderate class with a minimal AP class.*

The results associated with the first research question were mixed. On the one hand, authentic pedagogy was positively associated with student performance on the graduation exam based on the regression analysis. The analysis, however, that best addressed the research question, the one-way ANOVA, indicated that students in the sample who received higher levels of authentic pedagogy did not outperform their peers in the lower authentic pedagogy classes. Further study is needed to verify this finding since this analysis was limited to a small, carefully paired subsample of classes from the study.

### Compounding Effect of Authentic Pedagogy

We also examined whether performance on the AHSGE improved with repeated exposure to classroom experiences that require students to perform challenging intellectual tasks.
We addressed this question by looking specifically at students who experienced multiple courses at the highest level of authentic pedagogy exhibited by our sample of teachers (the moderate authentic pedagogy level). A prior moderate variable was created to designate students as having received 0, 1, or 2 history courses featuring moderate levels of authentic pedagogy. The one-way ANOVA used to investigate this research question included 328 students with no exposure to moderate pedagogy, 292 with one class of moderate authentic pedagogy, and 58 students in the group that experienced two moderate authentic pedagogy courses. Table 4 provides a breakdown of the results (first row of data). Repeated exposure to courses featuring moderate authentic pedagogy had a significant effect on student achievement on the graduation exam \((p < .001)\). Although the means were significantly different, the effect size was not very large \((\eta^2 = .04)\).

### Table 4

<table>
<thead>
<tr>
<th></th>
<th>No moderate AP classes</th>
<th>One moderate AP class</th>
<th>Two moderate AP classes</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (SD)</td>
<td>Mean (SD)</td>
<td>Mean (SD)</td>
<td></td>
</tr>
<tr>
<td>Advanced Placement</td>
<td>539.08 (70.257)</td>
<td>556.38 (65.686)</td>
<td>585.91 (47.924)</td>
<td>14.13***</td>
</tr>
<tr>
<td>Students included</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Advanced Placement</td>
<td>536.16 (70.263)</td>
<td>534.12 (64.369)</td>
<td>569.53 (51.762)</td>
<td>2.121</td>
</tr>
<tr>
<td>Students excluded</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. ***\(p < .001\)

Hochberg’s GT2 post-hoc comparisons of the three groups indicated the group that experienced two moderate authentic pedagogy classes had significantly higher scores on the graduation exam than students with only one moderate class \((p = .006)\). The two-class 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 did not experience any classes at the moderate level. The students with one authentic pedagogy course performed significantly better, \(p = .004\).

Each group in this analysis included regular and advanced placement history students. A larger percentage of advanced placement students, however, were in the 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, another one-way ANOVA was conducted that excluded the advanced placement students (second row of data in Table 4). 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.

**Discussion**

This study sought to determine the relationship between authentic pedagogy and student performance on the Alabama High School Graduation Exam (AHSGE). The results of this study suggest that the use of inquiry-based instruction does not force teachers to sacrifice student performance on high-stakes tests of lower-order knowledge. Students in this study who experienced higher levels of authentic pedagogy did as well as their peers in classes that featured lower levels of authentic pedagogy on the history portion of the AHSGE. They did not, however, experience the type of performance benefit that might have been expected based on the results of similar studies (Newmann, Bryk, & Nagaoka, 2001; Saye & SSIRC, 2013). Work by Newmann and his colleagues 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 (Newmann, Bryk, & Nagaoka, 2001). The researchers explained these results in terms of vocabulary acquisition and motivation to learn and 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 the study included such a small sample of teachers at the grade level the graduation exam was administered (N =4). When data from our study were incorporated into the larger SSIRC project, the results were more consistent with the work of Newmann, Bryk, and Nagaoka (2001) (Saye & SSIRC, 2013). Our 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 were needed to achieve a noticeable performance benefit for students on standardized tests of lower-order knowledge. 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 the previously mentioned theory regarding vocabulary acquisition and motivation doesn’t hold true for this particular test. This seems unlikely since the main SSIRC study found that students taught by teachers demonstrating higher levels of authentic pedagogy were more successful on state tests of this nature (Saye & SSIRC, 2013).

Our study also investigated the impact of the independent subcomponents of authentic pedagogy (tasks and instruction). The regression analysis indicated a statistically positive relationship between authentic instruction and student performance on the graduation exam. In the case of authentic tasks, however, the relationship was statistically significant, but negative. The discrepancy, once again, between these two variables could simply be due to the narrow focus of this study. It is also possible that some teachers adopted more challenging tasks designed by their peers or from other sources, without altering their usual instruction to any great extent. This was evident during the study on several occasions. A teacher, for example, classified in the minimal authentic pedagogy category submitted an illustrated timeline task that was designed by a teacher at the moderate level. The task scores for this activity were essentially the same for the two teachers, but the instruction score was much higher for the moderate teacher. The difference in scores on the instruction rubric was largely a result of the moderate teacher better anticipating the intellectual challenge associated with the activity and preparing
effective scaffolding accordingly. He set up the illustrated timeline activity with a thorough introduction to activate student prior knowledge and then guided students through multiple examples to model the brainstorming process and his expectations for the final product. The students in his class were required to complete the assignment in pairs to encourage discussion of ideas (not the case in the minimal authentic pedagogy class). They were more deliberative and thoughtful in completing the assignment compared to the students in the minimal authentic pedagogy class who completed the assignment at a much faster pace. Success on this assignment was largely a function of the teacher knowing how to structure the assignment to maximize higher order thinking. This is a skill that is developed over time as teachers implement inquiry-oriented activities.

It seems logical that challenging tasks, by themselves, would not provide a large 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.” The moderate authentic pedagogy teachers likely scored higher on the authentic instruction rubric because they more routinely implemented challenging authentic tasks 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. A unique aspect of our study, in comparison with previous AIW research, was the analysis of whether there was a compounding or increasing effect on student performance when students experienced multiple classes at the moderate authentic pedagogy level. We found 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, this finding is questionable since there were more advanced placement history students in the group that experienced two courses at the moderate authentic pedagogy level than in the other two comparison groups. When the advanced placement students were removed from a subsequent analysis, the students who experienced two courses at the moderate level still achieved higher scores than their peers, but the finding lacked significance. The results of this aspect of our study were clearly not as strong and conclusive as those identified in Klentschy’s research (2001).

Research has consistently shown that students are not likely to experience high levels of authentic pedagogy in their classes (Bryk, Nagaoka, & Newmann, 2000; Ladwig, Smith, Gore, Amosa, & Griffiths, 2007; Newmann & Associates, 1996; Saye & SSIRC, 2013). As a result, students who experience a course with higher levels of authentic pedagogy may have to adapt more than usual to different expectations (i.e. constructing knowledge vs. simply replicating it). How long does it take the average history student to truly reap the benefits of authentic pedagogy? What outcomes should be reasonably expected of students taking their first course that features moderate to substantial authentic pedagogy? The investment of time and resources to prepare teachers to provide higher levels of authentic pedagogy might be more widely embraced if policy makers could more readily answer these questions and discern the impact on learning when students experience a progression of authentic courses through their junior high and high school history courses.

Teachers and administrators have long been concerned that the focus on depth required by inquiry-based teaching will result in the coverage of fewer topics and thus leave students ill-prepared for high stakes exams of lower-order knowledge (Newmann, 1988). These concerns may become exacerbated as more systems consider value-added models of teacher assessment.
The results from this study and the larger SSIRC study (2013) indicate no disadvantage in performance on high stakes tests for students who experience moderate to substantial levels of authentic pedagogy. Recent work by Parker et al. (2011) investigating the impact of a redesigned, project-based AP Government and Politics curriculum also support this conclusion. These findings and similar studies involving “ambitious” teachers who promote higher order objectives in their classroom despite pressures from high stakes tests (Grant & Gradwell, 2010) provide a basis for teachers and administrators to engage students in more meaningful research-based learning. Furthermore, authentic intellectual work standards promote the literacies reflected in the Common Core Standards for History. As Common Core assessments are implemented in the schools, authentic pedagogy practices may prove critical to promoting student success on these more demanding measures of proficiency.

Conclusion

This study addressed a need in the field by examining the impact of authentic intellectual work on student achievement on a high stakes test of history knowledge. The results suggest that authentic pedagogy did have a small, positive association with student acquisition and retention of lower-order knowledge, but not to the extent demonstrated by the most relevant AIW study (Newmann, Bryk, & Nagaoka, 2001) or the larger SSIRC study (Saye & SSIRC, 2013). An element of this study that deserves additional research is the “compounding effect” of authentic pedagogy. This study does not provide a definitive answer to the question of whether student achievement improves with multiple courses at the moderate authentic pedagogy level. Longitudinal studies, with larger samples 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. The effort seems worthwhile if the trend identified in this study is confirmed.

Author Note

This research is part of the Social Studies Inquiry Research Collaborative project focused on analyzing the relationship between authentic intellectual challenge and student achievement. See Web-Based Reference for website identifying members of the project.

References


(Ed.), *Measuring history: Cases of high-stakes testing across the U.S.* (pp. 195-220). Greenwich, CT: Information Age.


**Web-based References**


Members of the project are listed on the following site: [http://www.auburn.edu/academic/societies/ssirc/member.html](http://www.auburn.edu/academic/societies/ssirc/member.html)

**Author Bios**

**Lamont Maddox** is an Assistant Professor at the University of North Alabama. His research interests include authentic intellectual work, problem-based instruction, assessment in social studies, and the development of historical thinking skills in students. Email: lmaddox@una.edu

**John Saye** is an Alumni Professor of Secondary Social Science Education at Auburn University. His research interests include authentic pedagogy, technology-enhanced learning environments, and teacher thinking.