MEASURING THE IMPACT OF THE DOCUMENT-BASED LESSON CYCLE ON EIGHTH GRADERS' ABILITY TO ANALYZE HISTORICAL DOCUMENTS

A Record of Study

by

BYRON RAY WILLIAMS JR.

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF EDUCATION

Chair of Committee,	Lynn M. Burlbaw
Committee Members,	Larry Kelly
	Sharon Matthews
	Anthony Stranges
Head of Department,	Michael De Miranda

December 2019

Major Subject Curriculum and Instruction

ABSTRACT

As a discipline, history sometimes gets a bad reputation because many students believe that it is all about memorizing long lists of dates, disputes, and dead people's names. Too often history teachers reinforce this belief by emphasizing historical content over the reasoning skills and disciplinary practices that historians engage in daily and that are transferrable from one social studies class or text to another. When students have opportunities to practice historical thinking in class, instead of just memorization, they encounter a variety of perspectives in different sources and learn to appreciate some of the nuance, complexity, and ambiguity in historical texts. In this study, two 8th grade history teachers at an independent school integrated the Document-Based Lesson Cycle into half of their classes during a six-week intervention to determine if students in the experimental group would outperform students in the control group on assessments of content knowledge and historical reasoning skills. Most of the Document-Based Lessons were developed by the Stanford History Education Group and are available on its website. Modeling of historical thinking by the teacher, analysis of multiple sources in each class period, and discussions about procedural knowledge are key components of every Document-Based Lesson. The control group had lessons that covered similar content, but the teacher-led discussions focused on content rather than on skills and most of the sources were secondary instead of primary.

Students took a four-part pretest and a four-part posttest that assessed both their content knowledge and skills proficiency. By the end of the intervention, there was no

statistically significant difference between the groups in terms of their content knowledge. The three skills assessments also revealed no statistically significant difference between the groups. However, the experimental group did demonstrate a greater ability to generate plausible historical claims that are supported by multiple historical documents. There were also some qualitative differences between the two groups in their discussions and on their writing assessments.

DEDICATION

"Let's face it: history is property; it is currency. Among other uses, we claim it to shape identities. We appropriate it to speak to contemporary political circumstances and to make our own history. And, we just plain use it to feed our curiosity. But to state the obvious and the too often unstated, the subjective experience informs public, political activism at every turn."

--Chana Kai Lee

For Gabrielle and Olivia

ACKNOWLEDGEMENTS

Records of study do not write themselves and as lonely as the process can be at times, this particular study was truly a team effort. At every step along the way, people came alongside me to provide encouragement, guidance, support, expertise, and advice. I am so grateful for my committee chair Dr. Burlbaw who mentored me and advocated for my students, for common sense, and for quality history education. The professionalism, intellectual curiosity, and encouragement of my other committee members, Dr. Kelly, Dr. Matthews, and Dr. Stranges inspired me.

When I was on the brink of giving up, Dr. Viruru and Ambyr Rios provided wise counsel and simply refused to let me quit. I want to thank them and all of the other teachers and students that I had the opportunity to work with at Texas A&M. I also need to thank Dr. Hutchins for helping me to make sense of the data I collected and Sharon Smith for her careful editing and help with APA formatting. My school's administrators and teachers supported my journey through this doctoral program.

My daughters did not read or write a single line of this study, but they inspired me to keep writing and put up with me spending way too much time sitting in front of a computer screen. They also kept me sane by encouraging me to join them on walks to the park or on the couch for movie night. I am grateful for my parents for convincing me that I could complete this type of study. I thank my wife who has had to listen to me talk about historical reasoning skills and DBQs for over two years. She has been my biggest cheerleader, coach, counselor, and friend.

TABLE OF CONTENTS

LIST OF FIGURES

Page

LIST OF TABLES

Page

CHAPTER I

INTRODUCTION

National Context

Historian Chana Kai Lee (2001) argued that history is the cultural and emotional currency that we use to construct our identities, public policies, TV and movie narratives, and political arguments. History is not only a discipline that documents the debates of the past and encourages students to debate the present, but it is also a subject that has become the center of heated debates on radio talk shows, on cable news networks, and in legislative halls around the country. Politicians and policymakers debate the role of history in primary and secondary schools. Historians also disagree about the meaning and purpose of history and they argue about how history should be taught and tested. The growing consensus among historians and history educators is that the study of history needs to focus more on skills and less on mere rote memorization of disconnected facts.

The debates about what should be taught in history classes and how those classes should be taught are not new. In 1989, the Bradley Commission, made up of over a dozen history professors and respected secondary history teachers, organized to evaluate the state of history education in the United States, concluded, like many other history educators, that history instruction was in peril. The commission noted that in 1989, fifteen percent of U.S. students took no history classes in high school and nearly 50% took no Western Civilization or World History course. These educators agreed that history should play a prominent role in secondary education and argued that its disciplinary practices are the best hope of equipping students to

cope with change (Bradley Commission, 1989). Since 1989, many history educators have attempted to define historical thinking and to identify its components.

However, even the Bradley Commission represents a relatively recent critique of history education. Truman L. Kelley (1914) of Columbia Teachers College published his *Education Guidance: An Experimental Study in the Analysis and Prediction Ability of High School Students*. In this work, he operates under the assumption that one way to measure historical ability is to evaluate students' ability to analyze a historical narrative and accurately predict the next event or most likely outcome. Kelley claimed, based on the poor performance of students on the assessment, that traditional methods of teaching history were insufficient to prepare students to think critically about historical issues. Three years later, in July of 1917, B.R. Buckingham, who served as the Director of the Bureau of Educational Research at Ohio State University, presented at the fifth annual meeting of the National Education Association of the United States and acknowledged that many historians and history educators viewed mere recall of historical information as an insufficient measure of a student's historical thinking ability. However, Buckingham administered tests to 8th-grade students and found that there was little to no difference between students' ability to recall and analytical questions. Bell and McCollum (1919) conducted a study in 1919 that measured students' historical knowledge. They chose to focus on content rather than skills or higher level thinking because they were influenced by Buckingham's argument that there is little need to distinguish between content and thinking skills in history assessments since they yield similar results. J. Carleton Bell of the Brooklyn Training School for Teachers and President of the Llano County School Board of Trustees D. F. McCollum bemoaned the fact that students were only able to answer 33-35% of their basic recall questions correctly, and they concluded that poor instruction was to blame. All of these early

twentieth-century assessments of historical knowledge and skills revealed a concern about the effectiveness of history instruction in the United States. Students were not able to obtain and retain the most basic levels of historical knowledge. Though there was some disagreement about how to most effectively and efficiently assess historical thinking or ability, there was and still is agreement that rote memorization of disparate facts was not producing the desired results in terms of acquisition of factual knowledge or practical skills.

Sam Wineburg, a professor at Stanford University and a pillar in the field of history education and the assessment of historical thinking, notes in his 2018 book, *Why Learn History (When it's Already on Your Phone)?*, that not much has changed in history instruction or assessment since World War I. There have been commissions, such as the American Historical Association's 1930 *Commission on the Direction of the Investigation of History and Other Social Studies in the Schools,* which reached the same conclusions as Kelley, Buckingham, Bell, and McCollum. There have been conferences, consortiums, panels, articles, and books all pointing toward a more skills-based approach to history instruction and assessment, but the lectures, the marches through heavy textbooks, and the pages and pages of multiple-choice questions persist. The discipline that is best suited to teach students to avoid repeating the same mistakes is guilty of repeating the same mistakes for at least the last century. At this point, it suffices to observe that history education experts agree that history education in the United States needs improvement.

Situational Context

Independence Academy, located in a heavily wooded and highly affluent suburb in the Southwest, has long enjoyed a reputation for academic excellence and strong leadership. Four years after welcoming its first class to learn around the founder's kitchen table and just before

the start of World War I, Independence Academy had grown from one class of kindergartners to three grade levels and thirty students. After nearly three decades of continuous growth, the school continued to expand, adding a high school division in the heart of the Great Depression. Shortly after the founder's retirement in the mid-1900s, the school seized an opportunity to relocate to a 40-acre plot of land owned by a member of the school's board of trustees. The third headmaster of Independence Academy began his tenure in the late 1970s and continued the school's development projects. His goal was to ensure that the school became one of the most elite and successful in the country. He resolutely believed in the need for strong curriculum and excellent faculty, but he also believed that state-of-the-art facilities were essential to success. To help alleviate overcrowding and to give both middle and upper school students and faculty their own space, he led the effort to construct the middle school building; I have the privilege of teaching in this building to this day. Of all his projects, the construction of the middle school for fifth through eighth-grade students was one that made him very proud. Every year, as eighth-grade dean, I have the honor of announcing the winner of a leadership award to the eighth-grade student who has contributed the most to the school community during his or her time in middle school.

During orientation at Independence Academy, I was told that members of the school community referred to it as an independent school, rather than a private school, and I was encouraged to do the same. Independent schools are private, but not all private schools are independent. Many private schools are started, managed, and governed by a religious, corporate, or non-profit institution; Independence Academy is not. The school's independent spirit is reflected in the fiercely independent faculty who often resist attempts at institutional standardization. When I began teaching at this school in 2011, after five years working in a

public school district, I noticed that the Independence Academy did not use the state's skills or content objectives. It later became clear that the history department, in particular, had many documents produced by individuals and groups of teachers that identified, outlined, and described different skills that students were supposed to learn at different grade levels and in different subjects. Some of these documents were in binders in the offices of administrators and in teachers' classrooms and others existed in various online repositories.

As a new teacher to the school, I wanted to get my hands or eyes on a clear scope and sequence and a list of skills objectives. Though several efforts to map these skills had taken place before my arrival and were in the works during my first couple of years at the school, our history department is just beginning to embrace and utilize national standards to inform planning and instruction. The Lower School or elementary division of Independence Academy, which includes Pre-K through 4th Grade, has outlined the social studies skills and content that are taught at each grade level and continues to build on it each year. The Social Studies Workgroup in the Lower School measured each grade level's social studies curriculum against the state standards and several sets of national standards. The Upper School (grades 9-12) has many A.P. classes, so they have the skills and content objectives produced by the College Board. The other high school and middle school classes each have course syllabi, but there is not yet a systematic way to track the skills progression of individuals and classes of students from year to year.

Identifying the Problem

Without a clear scope and sequence and a vertically-aligned progression of skills objectives, teachers at the school must rely upon their own expertise as historians and educators to determine when and how to teach specific historical thinking and reasoning skills. Teachers have not adopted a single agreed-upon method of assessing historical thinking skills aside from

the research papers that are required every year starting in sixth grade. Teachers in the English and History Departments worked together to develop guidelines for research papers that emphasized thesis statements, evidence, analysis, and paragraph structure. As a part of a campus-wide curriculum mapping effort in 2010 and a departmental effort again in 2016, teachers in the Middle School History Department did create digital copies of the scope and sequence for each course and identified the skills that were taught in each unit of study (R. Crockett, November 2, 2018, personal communication & K. Casey, November 13, 2018, personal communication). Students begin taking history tests in fifth grade, but the design of these tests can vary quite a bit from one grade to the next. There have been efforts to increase the horizontal alignment of assessments in order to ensure fairness and equity for students who are taking the same course from different instructors (T. Brown, personal communication, November 6, 2018). As recently as the spring of 2019, the Middle School History Department adopted a national set of standards and has plans to identify skills and content gaps in the curriculum over the course of the 2019-2020 school year.

One of my biggest desires, as an eighth-grade teacher, is to see my former students thrive in our upper school history classes. As one of two eighth grade history teachers, I am responsible for recommending which history course my students should take when they enter Upper School. Teachers of both the World History Advanced Placement and Global Studies courses have noticed gaps in the students' skills and content knowledge (Independence Academy Ninth Grade History Team, November 12, 2018, personal communication). The purpose of this study is to begin to develop a system for teaching and assessing the skills that will be most relevant in upper school history classes.

Relevant History of the Problem

In 2013, Independence Academy introduced a two-year World History Advanced Placement (WHAP) course for sophomores and freshmen. This course was introduced because ninth graders were not required to take history and because offering an A.P./honors course would give a GPA bump as an incentive to enroll; advanced courses are more heavily weighted than regular courses. The school's accrediting organization recommended that all ninth graders be required to take history (B. Wilson, December 17, 2018, personal communication). This recommendation not only led to the introduction of the WHAP course, but it also inspired the Upper School History Department to develop an on-grade-level course for ninth graders. This Global Studies course was introduced in the 2017-2018 school year to help bolster the historical reasoning and study skills of the students who did not qualify for the WHAP course. I would like to identify a set of skills that we teach and assess in all of our history classes and begin tracking varying levels of proficiency in the use of those skills. This would give our department a common language and shared expectations that we could use to hold students and teachers accountable.

Significance of the Problem

Several changes in the school and in the field of history education have sparked interest in better preparing students for high school AP history classes. Considering that, according to Upper School Head Tyler Brown (Email communication, January 2019), more than 60% of students at Independence Academy will enroll in at least one AP history class before they graduate, effectively teaching historical reasoning skills will make their transition to upper school smoother. The AP courses and assessments have changed in the past five years. They now assess less rote memorization and more complex cognitive processes. The 2019 version of

the test will not include Ancient and Medieval history (Strauss, 2018a). These were the eras covered in the ninth grade portion of our WHAP course. One of the biggest reasons given for this change is that it was extremely difficult for teachers to focus on skills and disciplinary practices when they were required to hurry students through the entirety of human history. This trend is consistent with the advice of historians and history education experts.

Eighth grade is an excellent place for our school to begin the process of identifying, assessing, and tracking historical reasoning skills because there is no pressure to rush through content to prepare students for an A.P. exam. We have the freedom to sacrifice content, if necessary, for the sake of skills instruction and practice. The skills are relevant in every history class. Susan De La Paz, Daniel Wissinger, and Laura Yee of the University of Maryland conducted a study with Ralph Ferretti and Charles MacArthur from the University of Delaware, (2012) and found that eighth-grade students were able to engage in disciplinary thinking and use multiple historical documents to construct argumentative essays. The eighth graders' arguments and essays were less sophisticated than the eleventh grade participants in the study, but the researchers saw this as evidence "that writers at both earlier and later secondary grades (and to some extent, good and poor writers within each grade) show similar patterns that may be seen as entry levels of performance to build on during history instruction" (De La Paz, Ferretti, Wissinger, Yee & MacArthur, 2012, p. 449).

Research Questions

These are the questions that this study attempts to answer in the context of eighth-grade classes at the Independence Academy School.

1. How effective was the Document Based Lesson Cycle at improving eighth grade Independence Academy students' ability to analyze primary and secondary documents on history assessments?
2. What is the relationship between explicit instruction on document analysis in eighth-grade history classes and achievement on assessments of document analysis skills?

Personal Context

Researcher's Role and Personal History

I began my career in education as 2003 Teach For America (TFA) corps member at Burbank Middle School in Houston Independent School District (HISD). I graduated from Xavier University of Louisiana with a degree in History and went through HISD's Alternative Certification Program. My coursework in education began with TFA and continued at the University of Houston and St. Thomas University. Most of my beliefs about history pedagogy came from my mentors at Burbank Middle School, Rice Pre-AP workshops, Teaching American History Summer Institutes at the University of Houston, and ongoing training within the school district. As much as I wanted to inspire students to love history and to sharpen their social studies skills, I felt that I owed it to them to make sure that they were adequately prepared for the Texas Assessment of Knowledge and Skills (TAKS). To do that, I became very familiar with the TEKS and frequently attended professional development sessions dedicated to interpreting and teaching them.

My experience teaching Gifted and Talented and regular classes at Burbank gave me many opportunities to practice differentiating instruction, re-teaching, tutoring, and assessing student performance. What I was supposed to teach, in terms of skills and content, was clear. My job was to figure out how best to teach it to the students in my classes. In 2011, I obtained a

master's degree from Southwestern Baptist Theological Seminary and began working at Independence Academy. I quickly learned that many of the school's teachers and departments developed their own course syllabi with little to no consideration of state or national standards or objectives. As the product and former employee of public schools, I did not initially know how to handle the autonomy of the independent school culture.

In 2015, my fifth year at Independence Academy, two of the other history teachers and I redesigned the eighth-grade history course. Instead of simply going through all of U.S. history chronologically, we decided to organize the course around four themes: 1. Migration, 2. The Role of Government, 3. War & Diplomacy, and 4. American Identity. We dedicated each quarter to the chronological exploration of one of those themes. We chose this approach because we noticed that students were having a difficult time identifying patterns of continuity and change over time and because we felt like the long lists of people, places, events, and dates were not in line with our beliefs about the goals of history education. In the process of developing this new curriculum, we used lesson and unit templates from *Understanding By Design* and identified enduring understandings, essential questions, and key terms for each theme, unit, and lesson (Wiggins & McTighe, 2005). We did not identify the historical reasoning skills that we planned to teach.

Journey to the problem

I enrolled in the Ed.D. program at Texas A&M to become a better history teacher and curricular leader. During my coursework in the Teaching, Learning and Culture Department, I read about how novices and experts construct and organize information differently in an article by Robert Glaser the founding director of the Learning Research and Development Center at the University of Pittsburgh (1984). This caused me to reflect on my own teaching practice and the

importance of developing disciplinary thinking and skills. Since I arrived at the school, my colleagues in the history department have emphasized the significance of skills, particularly in middle school where there is less external pressure to cover vast amounts of content. As I started researching disciplinary thinking, I first studied to the *A.P. United States History Course and Exam Overview* (College Board, *2017)*. The 2017 version of this document identifies four historical reasoning skills (Contextualization, Causation, Continuity and Change over time, and Comparison) and two disciplinary practices (Analysis and Argumentation). Since these skills extend across all A.P. History classes and our school has not adopted a set of vertical skills objectives means that we could use the A.P. framework to craft a system for teaching, assessing, and tracking students historical reasoning skills.

During my first internship, my field supervisor, Interim Head of School Payton Stiles recommended that I speak with curricular and administrative leaders in all three divisions to get a better picture of how they developed, evaluated, and implemented curriculum (October 25, 2018, personal communication). This series of meetings gave me a clearer picture of the work that has already been done to vertically and horizontally align the history curriculum. In my conversation with the chair of the History Department (or curriculum committee) in the Lower School, which includes grades Pre-K through 4, I learned that the Lower School Head formed departments and allowed teachers to sign up for the one that interested them the most. She challenged the social studies department to look at national and local social studies standards and to identify the ones they had already incorporated into their curriculum. They were also asked to see if there were any glaring gaps. They noticed that each grade level in Lower School already taught multiple lessons and units that tied into the geography, government, and economic social studies strands. The next phase of their curriculum development will seek to more intentionally

incorporate other strands such as culture, time, and civics (Q. Gomez, November 1, 2018, personal communication).

The work on the recently developed curriculum committees is an extension of the earlier curriculum mapping that took place from 2010-2012. During the curriculum mapping effort, the school invested in software that allowed teachers to create a digital scope and sequence for each course and to upload and share skills objectives, instructional materials, and assessments. The professional development for the curriculum mapping was led by divisional and departmental leaders and Heidi Hayes Jacobs, president of the Curriculum Mapping Institute and author of *Mapping the Big Picture: Integrating Curriculum and Assessment K-12* (1997). While teachers in all three divisions completed these curriculum maps, they continue to be more important planning tools for the leadership and instructional teams of the Lower School than they are for those in other divisions. While Social Studies is an increasingly important aspect of the Lower School curriculum, it is not a separate course. This is frequently the case in elementary schools (Lazer, 2015). The Lower School students do not receive any letter or number grades on their report cards. Their report cards have proficiency levels for dozens of grade-appropriate behavioral and academic categories such as math, reading, art, Spanish, and music. Social studies knowledge and skills are only assessed when they are embedded in other skills like literacy or math (K. Casey, November 13, 2018, personal communication).

Starting in fifth-grade, students begin taking U.S. History as a separate class. This is when students take their first history quizzes and tests and when they purchase their first history textbooks. They write multi-page research papers every year starting in sixth grade. Test and quiz items range from multiple-choice questions and matching to paragraph-length responses to writing prompts. Some teachers utilize project-based models of instruction while others

emphasize student-centered discussions. All of the middle school courses use a mixture of primary and secondary documents, but the department is still working to develop a shared language, a reliable and user-friendly tracking system, and assessment tools for the skills taught in middle school (R. Crockett, November 2, 2018, personal communication).

Nearly 70% of upper school students at Independence Academy enroll in at least one AP history course during their four years in high school. As an independent school that has not adopted the Common Core Standards or the Texas Essential Knowledge and Skills (TEKS), Teachers and departments have some freedom to design their curriculum in a way that they feel best meets their students' needs. However, in the past five years, the school has established a curriculum committee made up of the Department Chairs from each of the three divisions and the Interim Head of School (formerly the assistant head of school). Division Heads are welcome at Curriculum Committee meetings and are occasionally invited to attend to discuss specific issues (P. Stiles, October 25, 2018, personal communication). Part of the purpose of the Curriculum Committee is to evaluate the overall academic program of the school, set the course for its future, and increase the vertical and horizontal alignment of the school's curriculum. The commitment of the upper school to AP history courses, however, ties the teachers and the school to the College Board's expectations for a history course. The College Board's guidelines are important to the structure and curriculum of AP courses in the Upper School, but administrators, departments, and grade-level teaching teams also influence what is actually taught in Upper School classes. Upper School Head Tyler Brown (November 6, 2018, personal communication) informed me that teachers who teach the same subject at the same grade level must work together to develop instructional units and assessments. While the lessons and assessments do

not need to be identical, Brown explained that a relatively consistent experience between different sections of the same course was a matter of equity and fairness.

After meeting with leaders in all three divisions, I met with the four teachers who teach ninth-grade history, and they helped me to narrow the focus of my study. Prior to the meeting with these teachers, my plan was to focus on contextualization or causation because I had been warned that even our seniors struggle with identifying and explaining patterns of continuity and change over time. I looked at the AP results of some of our sophomores, juniors, and seniors and did not notice that one particular skill stood out as particularly weak when compared with the others. The ninth grade teachers explained that they do not explicitly emphasize the AP historical reasoning skills until the second half of the course because they spend so much of their time emphasizing the importance of time and chronological reasoning. They also dedicate a large portion of their time to teaching and modeling close-reading, note-taking, and evaluating the quality of various sources and types of evidence (Independence Academy Ninth Grade History Team, November 12, 2018, personal communication). A takeaway from that meeting was that our time in 8th grade would be better spent focusing on the AP disciplinary practices of analysis and argumentation rather than attempting to teach the loftier reasoning skills. Considering that most argument development in history classes requires students to first analyze at least one source, we agreed that document analysis would be the best starting point for my research.

Significant stakeholders

The 118 eighth grade students in our history classes represent the target population for our intended intervention and they are the most significant stakeholders in this study. Their skills are the ones that will be assessed, tracked, and hopefully improved as a result of this

process. The other eighth-grade history teacher, Mia Jones and I are the two teachers most directly impacted by this intervention because we were responsible for implementing the new instructional approach and for assessing student performance on those assessments.

I was able to present my initial plans for this record of study at a middle school History Department meeting in early December 2018 to solicit feedback on the direction of the study. One of the results of that meeting was that the Department Chair scheduled a full-day meeting in January 2019 to discuss skills objectives and rubrics (R. Crockett, November 2, 2018, personal communication). This study could provide a model for skills instruction and assessment on our campus. We could also use the assessments and data to increase the productivity of interdivisional conversations. Ninth grade teachers would have more concrete performance data at the beginning of the year and they could use that data to identify trends that continue or change from year to year.

Important Terms

Division Head--Independence Academy uses this term to refer to the principals of its lower, middle, and upper schools.

Document Based Questions (DBQs)--A type of assessment used on Advanced Placement History Exams and in many secondary history classes. They usually require students to read through several primary documents related to a specific historical event or movement, develop an original thesis in response to a question about the documents and the time period, and then write a paragraph or short essay that uses evidence from the sources to support the historical claim (VanSledright, 2015).

History Assessments of Thinking (HATs)--Short assessment or test items designed to elicit the use of specific cognitive processes. They usually require students to analyze a primary

or secondary source and evaluate its reliability, usefulness, or purpose (Wineburg, Smith, and Breakstone, 2018).

Head of School--The chief executive officer for a school. Schuermann & McGovern (2016) describe the head of school at independent schools like Independence Academy as the person who "is ultimately responsible for managing the staff and all day-to-day operations of the school," while "the board hires and evaluates the head and makes certain that the school is fulfilling its mission and stewarding its resources wisely" (p. 569).

Historical Reasoning Skills--This is the term that the College Board uses for critical thinking skills that are specific to the discipline of history (College Board, 2017). They are sometimes called historical thinking skills.

Independent Schools-- "Private, nonpublic schools that are governed by independent boards of trustees" (Schuermann & McGovern, 2016, p. 569).

Lower School--Independence Academy and many other independent schools use this term to refer to elementary school. At Independence Academy, the Lower School consists of grades Pre-K through 4.

Upper School--The school uses this term to refer to its high school which consists of grades 9-12.

Think-Aloud Protocols (TAPs)--Interviews that are conducted during or shortly after an assessment that allows the researcher to hear from test taker's cognitive processes for analyzing and answering test questions (Ercikan, K., Seixas, P., Lyons-Thomas, J., & Gibson, L., 2015).

Weighted Multiple Choice Questions (WMCs)--Questions that assign higher point values to answer choices that represent more advanced historical reasoning skills and lower values to less advanced answers (VanSledright, 2015).

CHAPTER II

REVIEW OF SUPPORTING LITERATURE

The History Wars

In 1999, Sam Wineburg, in his popular *Historical Thinking and Other Unnatural Acts* did not go so far as to deliver the eulogy at the funeral of history education in the United States, but this history educator did claim that the discipline was undervalued, underappreciated, and most importantly under-taught. As recently as 2018, Jill Lepore, bestselling author and Harvard historian, notes that despite its poor standing in American schools, history curriculum continues to be a source of conflict between those on both sides of the political spectrum. Many see the history classroom as the place to advance their political agendas and ideologies, rather than a learning lab where students can sharpen their reasoning skills, empathize and argue with historical figures, and shape their own opinions and belief systems (Lepore, 2018).

The 2014 A.P. United States History Curriculum (APUSH) came under intense scrutiny and criticism because it was perceived by many conservatives to be an anti-American, liberal document designed to encourage history educators to present a harshly critical view of the United States. Independent scholar and educational consultant Jeremy A. Stern (2014) who earned his Ph.D. in history from Princeton, rejects the conspiracy theorist's view of the APUSH curriculum while acknowledging that objective observers can identify some legitimate political grievances with the curriculum. The College Board was in the news again in 2018 for changes it made to the World History A.P. course. Complaints from history educators that the new curriculum was too Eurocentric compelled the College Board to revamp the course curriculum

just months after releasing it (Strauss, 2018b). The socially and politically charged debates over what parts of the past should be taught are reminders that history is a powerful tool.

Historical Thinking--What is it? Why does it matter?

Many history educators have attempted to define historical thinking and to identify its components. The Bradley Commission (1989), a group of 17 scholars and classroom teachers, met to address the inadequacies of history education in higher education and in primary and secondary schools. Less than a decade after the Bradley Commission, the Organization of American Historians (1995) met and identified five standards in historical thinking: chronological thinking, historical comprehension, historical analysis, historical research capabilities, and historical analysis/decision-making. This meeting was not without conflict and criticism. Many resisted the conclusions and suggestions of the OAH (Appleby, 1995). Tim Huijgen who teaches history education at the University of Groningen in the Netherlands, Carla van Boxtel of the University of Amsterdam, along with Wim van de Grift and Paul Holthius, both of the University of Groningen (2017) add historical perspective taking to the list of historical reasoning skills and they divide this skill into three subskills: awareness of the danger of presentism, historical empathy, and historical reconstruction. According to Huijgen, van Boxtel, van de Grift, and Holthius (2017) the ability to contextualize and empathize with historical figures can prevent the unnecessarily harsh assessment or the beatification of historical figures.

Historical thinking skills are the cognitive processes that historians use to analyze and evaluate sources and to reconstruct meaningful narratives of the past. Historical thinking is not simply critical thinking applied to historical documents. As Jon A. Levisohn from the Department of Near Eastern and Judaic Studies at Brandeis University (2017) explains,

"Engaging in the work of any of these fields entails not the application of a generic disposition called 'critical thinking,' but rather the development of sensitivity to these challenges, familiarity with these moves, and appreciation of these traditions" (p. 625). Critical thinking, continues Levisohn, is similar to studying linguistics or language in general, while historical thinking is more similar to learning a specific language. The skills used to solve a complex physics problem are not the same ones used to evaluate the credibility of a secondary source. These disciplinary skills often go neglected because many teachers feel that they do not have time to identify primary documents that are appropriately challenging for their students and when they do find the sources they lack the time to effectively teach the reasoning skills to their students while still covering the long-list of content objectives required by their states and districts (A. Reisman, 2012).

Some teachers avoid teaching historical thinking because of the difficulty in grading the assignments, but high school educator Daniel Cowgill and University of Central Florida professor Scott Waring (2017) found that many history teachers fail to teach historical thinking because they lack the necessary skills and training themselves. Cowgill and Waring (2017) chose to compare the historical thinking skills of students and their teachers. They found that teachers and students made similar mistakes in historical thinking, implying that one reason why students fail to develop strong skills in history is that their teachers lack the same skills. They recommend that schools of education increase their requirements for discipline-specific coursework and that ongoing professional development provides more opportunities for teachers to develop and learn to teach and assess these skills. Wineburg, Abby Reisman of the University of Pennsylvania, and Brad Fogo of San Francisco State University (2007) concluded that even

though many studies of history instruction and assessment are weak on empirical evidence and valid measures, the sheer volume of articles and studies on the topic bode well for the discipline.

Levisohn (2017) believes that Wineburg overstates his case that historical thinking is such an unnatural act that most students are incapable of engaging in it without extreme interventions by expert history educators. Historical thinking helps to humanize students or make them more fully human, so, according to Levisohn (2017), it cannot be that unnatural. To support the critiques of Wineburg's claims and research, Levisohn reanalyzed Wineburg's Thinking Aloud interviews with students and attempted to show that some of the reasoning given as examples of failed historical thinking was actually rational and not necessarily a sign of students' inability to think like historians. Even if Wineburg overreaches in his arguments about how unnatural historical reasoning is to the novice, there is still a wide gap between the skill sets and knowledge bases that history teachers desire for their students and the ones they actually observe in the classroom and on formative and summative assessments. Historical thinking, if not unnatural, is at least cognitively demanding. Stéphane Lévesque, author of *Thinking Historically: Educating Students for the Twenty-First Century* (2008) and associate professor of History Education at the University of Ottawa, defines this type of thinking as "the intellectual process through which an individual masters--and ultimately appropriates--the concepts and knowledge of history and critically applies such concepts and knowledge in the resolution of contemporary and historical issues is extremely demanding and complex" (p.27).

Disciplinary Literacy

Historians read differently than other people. Though most secondary students will not pursue careers as professional historians, Reisman of the University of Pennsylvania argues that one job of the history teacher is to teach students to read like historians (Reisman, 2012).

Reisman makes this claim in part because of a widespread push in education to have literacy instruction embedded within each discipline. History is not the only discipline in which experts bring a unique set of skills, organizational structures, and epistemological beliefs to the task of reading. Mathematicians pay special attention to very precise meanings of words and phrases, while scientists and science teachers are more concerned about visual representations of textual data and the relationship between the visual and textual representations of information (Shanahan & Shanahan, 2008). Early reading skills, such as decoding and recognition of high-frequency words, are very generalizable to a wide variety of contexts. As students get older, their skills advance, and the texts they encounter increase in specificity, the skills needed to successfully interpret and use different sources become more discipline-specific (Shanahan & Shanahan, 2008).

Just as literacy skills become more specific as texts increase in difficulty and sophistication, the students' knowledge must become more discipline-specific. Experts possess two types of disciplinary knowledge: topic knowledge and domain knowledge (Alexander, 2003). Topic knowledge is an understanding of content in a particular field, whereas domain knowledge is knowledge of the organizational structures, heuristics, and cognitive tools needed to acquire, interpret, and use the disciplinary information. One key indicator that a student has begun the shift from basic to intermediate or competent thinking is that he or she is able to transfer knowledge to novel situations outside of the classroom (Frey, Fisher, & Hattie, 2016). Domain knowledge is not acquired through intellectual osmosis; sitting in a history class where students read historical texts and listen to lectures by an expert historian does not guarantee that a student will obtain historical reasoning skills. Domain knowledge is acquired in settings where teachers scaffold their lessons to make the information and skills accessible to their students and

when teachers model and give students opportunities to practice organizing information in ways that resemble the structures used by disciplinary experts (Levstik & Barton, 2015).

Three important heuristics used by historians are sourcing, corroboration, and contextualization (Nokes, Dole, & Hacker, 2007; Van Drie, & Van Boxtel, 2007; McCarty, 2016; Smith, 2017; Wineburg, 1990, 2010; Reisman, A. 2012). According to Wineburg (1990), though they do not guarantee successful and accurate historical comprehension, analysis, or interpretation, "These heuristics can be thought of as sense-making activities, for they help their user resolve contradictions, see patterns, and make distinctions among different types of evidence" (p. 77). Patricia Alexander, Distinguished Scholar-Teacher and Professor in the Department of Human Development at the University of Maryland, (2003), when describing the Model for Domain Learning, considered the use of multiple heuristics to be a significant benchmark in the journey from novice to expert. The difference between competent individuals and experts lies not only in the ability to use heuristics but in the ability of experts to contribute to new questions and theories to the field. Alexander (2003) explains, "to create new knowledge, experts must be well versed in the problems and methodologies of the domain and actively engaged in problem finding" (p. 11). Sourcing, contextualization, and corroboration are so important to the discipline of history because they allow experts to transfer their knowledge to new situations, documents, and issues (Frey, Fisher, & Hattie, 2016).

Some argue that one reason for the lack of historical reading and reasoning skills of history students is the lack of alignment between the disciplinary practices of expert historians and the instructional methods of history teachers (Nokes, 2011; Stoel, 2015). Nokes (2011) went so far as to claim, "History, as it has traditionally been taught, is unique among all other secondary subjects in the disparity between the behaviors of those who are in the field--

historians--and those who are in the classroom--history students" (p. 395). Calls for increasing the alignment between the work of historians and history students date back more than a century. The American Historical Association (1898) established the Committee of Seven to make recommendations about history education; the committee argued against the abandonment of the history textbook for a collection of excerpts from primary sources, claiming that such an approach encourages students to form opinions based on a limited and fragmented set of sources. The Committee of Seven (1898) also advocated for history courses that provided "limited contact with a limited body of materials" and "an examination of which may show the child the nature of the historical process, and at the same time may make the people and events of bygone times more real to him" (p. 104). In a report to the New England Teachers' Association, a special committee convened to research best practices for teaching history concluded that "some place should be found for the use of original material in the history courses of our secondary schools, because those courses may thereby be made more interesting" (New England History Teachers' Association, 1902, p. 8). Like the Committee of Seven, the New England Teachers' Association warned against prioritizing primary source material over well-written and organized textbooks.

Even though leaders in the field of historical thinking at the turn of both the twentieth and twenty-first centuries agreed that primary sources were valuable tools in a history classroom, there seems to be a trend toward greater distrust of textbooks. Jerry Moore (University of Virginia), James Alouf (Sweet Briar College), and Janie Needham of the Colorado State Department of Education (1984) placed part of the blame on history textbooks for the underdeveloped historical reasoning skills of secondary students claiming, "Traditional textbook history offers no understanding of how historical thinking proceeds" (p. 58). More recently,

University of Kentucky history education expert Linda Levstik and Keith Barton at Indiana University (2015) warned that textbooks encourage the very myopia that the discipline is intended to combat because "the range of interpretations traditionally found in textbooks and school curricula has been extremely small" (Chapter 1, "History is More Than Politics, para.1). Wineburg's (1999) complaint about history textbooks is that they approach history differently than historians do when writing for each other. History textbooks lack the uncertainty, contradiction, and complexity present in most historical writing. Harvard historian and bestselling author, Jill Lepore (2018), observed, "Some American history books fail to criticize the United States; others do nothing but" (p. xix). Heavy reliance on textbooks and lectures in history classes in secondary and college classrooms results in student misunderstanding of how history is constructed and an erroneous belief that history is "the straightforward acceptance or rejection of authoritative, binary stories presented to them as self-evident by teachers, textbooks authors, historians, or other authorities (Lévesque, 2008, p. 138).

Primary sources, unlike most history textbooks, broaden the perspectives investigated in secondary history classes and cultivate the development of historical reasoning skills and the retention of content knowledge. Jeffery D. Nokes at Brigham Young University and J.A. Dole and D.J. Hacker, both at the University of Utah, (2007) found that on assessments of content and historical reasoning skills students in classes that relied on multiple documents outperformed those who relied primarily on textbooks. The use of multiple documents does not guarantee that students will learn to use the heuristics that history educators claim are so important. Students need to be able to "reason about documents," which means to "evaluate each piece of information on the basis of the type of document it is" and to "reason with documents" or "to use document information when solving a problem" (Rouett, J., Britt, M. A., Mason, R. A., &

Perfetti, C. A., pp. 478-479). Reisman (2012) conducted a study in urban high school classes using the *Reading Like a Historian* (RLH) curriculum and found that the treatment group showed more growth than the control group in sourcing and close-reading skills. Some skills--contextualization and corroboration--did not improve in the treatment or control group. Nokes (2011) identified four major barriers to adolescents' ability to read like historians: cognitive complexity, misapplied background knowledge, unsophisticated worldviews, and a misunderstanding of the nature of history.

Testing and Measuring Historical Thinking Skills

Assessing historical thinking is a daunting task even for the most experienced test developer (VanSledright, 2015). Once the researchers, testers, or policymakers have narrowed the list of skills, the writers must begin the work of determining exactly how to measure proficiency in those skills (Reisman, 2015). In cases where there is a need to test a large number of students, the developers must weigh the benefits of more authentic assessments against the financial and time costs (Lazer, 2015). The easiest and most efficient way to assess historical thinking is to just use multiple-choice questions. All twenty-three states that test history at the state-level use multiple-choice questions. In fact, fourteen of those states use only multiple choice questions (Smith, 2017). The 2018 State of Texas Assessments of Academic Readiness Eighth Grade United States History exam consisted of 44 multiple choice questions. Most of the questions required students to analyze a quote, image, or chart and select the best answer from a list of four options. The 11th grade U.S. History STAAR exam used the same style and format but included 67 questions. With the increased emphasis on improving assessment and instruction in history classes, many scholars and test developers have been evaluating and changing the structure and style of the course outlines and the tests themselves. Bruce

VanSledright (2014), Professor Emeritus at the University of North Carolina Chapel Hill and a leading scholar in the field of assessing of historical thinking, claims that the multiple-choice questions so common on history assessments do not actually measure the skills historians, educational policymakers, and classroom history teachers value most. Multiple-choice questions are much better for measuring memorization and vocabulary than they are for measuring historical reasoning. In a collaborative study between the College Board and the Regents Research Fund, Pamela Kaliski, Kara Smith, and Kristen Huff (2015) found that students were frequently able to arrive at the correct answer without utilizing the skills that the items were designed to measure, while others used the correct cognitive process but still selected incorrect answers. This brought into question the effectiveness of the questions at assessing the target skills. The tests yielded results, but Kaliski, Smith, and Huff concluded that those results were misleading because they were based on invalid assessment tools.

One of the tools most frequently used to evaluate the validity of historical thinking skills questions is the thinking aloud interview. When using this strategy, researchers interview test-takers during or after an assessment, prompting them to explain what they are thinking as they are reading the questions and formulating or picking their answers (Kalisiki, Smith & Huff, 2015). Smith (2017) used 50-minute concurrent and retrospective thinking aloud interviews for all twelve students in the study to question them about their answers to multiple-choice questions. Smith found in the TAPs interviews that the Historical Thinking Test (HTT) elicited the intended cognitive processes more frequently than the multiple choice questions typically used on standardized history exams.

University of British Columbia professors Kadriye Ercikan, Peter Seixas, and Lindsay Gibson along with Columbia University professor Juliette Lyons-Thomas (2015) used an

Evidence Centered Design (ECD) to create a Historical Thinking Assessment Tool. They chose to only assess three dimensions of historical thinking so the assessment could be completed in one hour. They used concurrent Thinking Aloud Protocols (TAPs) to determine if the assessment items actually assessed the intended skills. The researchers compared verbalizations of the use of historical thinking from the TAPs to student performance data to see if students who verbalized use of historical thinking actually outperformed those who did not.

Using TAPs can give researchers an idea of how effective a test item was at eliciting the intended cognitive process, but TAPs have their weaknesses as well. Like all quantitative and qualitative data, the results of TAPs interviews must be interpreted and the interpretations of these interviews can vary from one researcher to the next. Wineburg (2001) and Levisohn (2017) looked at the transcript of the interview with a high school junior named Derek and reached contradicting conclusions about the student's ability to empathizing with the participants in the Battle of Lexington Green. In Wineburg's study, Derek read a series of documents about the Battle of Lexington Green and explained to the interviewer that the conflict seemed to be more of a disorganized skirmish than a systematically-executed battle. However, when asked to pick a painting that most accurately illustrated the conflict, Derek chose the one with troops in battle lines and colonists taking cover in fortified positions. The discrepancy between Derek's interpretation of the primary sources and his selection of a visual representation was used by Wineburg as evidence that he lacked the historical reasoning skills needed to properly interpret the sources. Levisohn disagreed and defended Derek's logic, claiming that his selections and descriptions represented age-appropriate analysis considering his understanding of military conflicts and the amount of information that he had access to at the time.

Gabriel Reich (2015), Associate Professor of Secondary History/Social Studies Education at Virginia Commonwealth University, who analyzed student performance on the New York Regents History Exam, argues that because they tend to measure the lowest forms of historical thinking and because they are incapable of measuring the depth of understanding in students' thinking, multiple-choice questions should be abandoned. Multiple-choice questions are in Reich's (2015) opinion, for bureaucrats who need statistical data, spreadsheets, and line graphs, but historians are attempting to develop a more nuanced view of history; their goals are not aligned.

Kalisiki, France, Huff, and Thurber (2011) presented on a study of seventeen World History AP students who were given fifteen multiple-choice questions, thirteen from the newly formatted exam and two from the old format. They also used TAPs interviews to see if the new questions were more effective at eliciting the intended reasoning skills. When coding their transcripts, Kalisiki, France, Huff, and Thurber identified several non-history specific tools and skills that students used to answer questions including guessing, process-of-elimination, scaffolding, and use of background knowledge. Since Kalisiki, France, Huff, and Thurber were presenting on behalf of the College Board which continues to use multiple-choice questions, they certainly were not arguing for the elimination of all multiple-choice questions. They argument support the creation of better multiple-choice questions.

Ercikan and Seixas (2015) encourage test developers to eliminate certain test items that rely on information that is not given in the exam itself because it puts students with limited background knowledge at a disadvantage. Their point is that it is difficult to know if a skill is actually being assessed if students could miss the question simply because they lack the requisite background knowledge. It is as if Ercikan and Seixas want to isolate the skills from the content.

Though the College Board, the Organization of American Historians, and the Bradley Commission all argued for prioritization of skills over rote memorization, they were not arguing for the abandonment of historical knowledge. Perhaps, Ercikan and Seixas just want to make sure that the historical reasoning items actually measure skills and not background knowledge or reading skills. They argue for an increase in the validity of the questions test developers ask history students.

Wineburg claims that history is an unnatural act; perhaps the same could be said of testing, especially the testing of historical reasoning skills. This would explain the difficulty in confirming the validity of test questions, student answers, and interpretations of the result. Johan Samuelsson and Joakim Wendell (2016) at Karlstad University in Sweden analyzed the historical thinking of Swedish students in the nation's new standards-based history curriculum, and they concluded that at least part of the challenge that students faced when attempting to use historical thinking was the result of the rigid and high-stakes nature of testing. Multiple-choice questions often test the easiest and least relevant aspects of historical thinking while essay and short answer questions give test developers and teachers more insight into student thinking, but these items are time-consuming to grade and require graders to interpret and evaluate the meaning, accuracy, and validity of the student responses (Lazer, 2015). The Swedish study mentioned above only analyzed student responses to two questions about one primary document. The students were asked to explain why a gravesite is a valid source about Viking life and culture, while a toy Viking ship is not, and then they were asked to explain the problem with using a gravesite as a historical source (Samuelsson & Wendell, 2016). While these questions require students to do work similar to that of an actual historian, different graders could reach different conclusions about the students' answers.

There is no one-size-fits-all model for assessing or testing historical thinking. Policy makers and district-level managers may need a different type of information about student progress than a classroom teacher, so there may be a need for students to engage in various types of assessment in order to provide different stakeholders with the information they need to make effective decisions. The classroom teacher, who is concerned primarily with the individual progress of the students in his or her classroom, might not benefit as much from standardized tests with multiple-choice questions (Reich, 2015). VanSledright (2015) recommends that teachers develop assessments—not necessarily tests—that emerge from the actual learning in the classroom. This gives the assessment what VanSledright calls ecological validity. The past is already unfamiliar to students, the format of the assessment does not need to add to the unfamiliarity.

Historical Thinking in 8th Grade History Classes

As an eighth-grade history teacher, I am aware that some of the challenges present in teaching and assessing historical thinking skills in a high school class are compounded when applied to middle school classrooms. Eighth graders typically have even less background knowledge than high schoolers, lower literacy levels, and less experience analyzing and evaluating primary and secondary documents. They also struggle more with abstract reasoning than older students. Despite the challenges, middle school students can engage in historical thinking. The Montgomery County Public Schools Office of Curriculum and Instruction and its Office of Shared Accountability commissioned a study that was designed to evaluate the effectiveness of the district's eighth grade advanced history curriculum at preparing students for high school history classes (Maina, 2015). The study tracked 31,253 students from 2009-2013. The study found that participation in the advanced program did increase the chances of students,

particularly African American and Hispanic students, enrolling in AP history courses. There was very little difference in the AP test scores and course grades of students who took the advanced eighth grade course and those who did not. However, the researchers warn against concluding that the advanced curriculum did not impact student performance. Maina (2015) suggests that students who chose to take the AP course without enrolling in the advanced eighth-grade course may have possessed enough academic promise, skill, and motivation to overcome any advantages gained by participation in the advanced course.

Like high school teachers, eighth-grade teachers should not lean exclusively on standardized summative assessments to measure students' historical thinking. Sergio Méndez and Felipe Tirado (2016) at the National Autonomous University of Mexico recommend using formative assessments in a standards-based curriculum to give students a continuous loop of feedback on their skill level and progress. In their study, teachers used a rubric with six procedural concepts (historical significance, consequences, effects in the present, causality, historical sources, and envisioning future events) and six skill levels for each. Eighth grade and high school teachers could learn from this model because it allows teachers and students to track not only skill attainment but also skill progression. Teachers could provide more accurate and timely data about where their students are on the spectrum of historical thinking ability and allow teachers to identify relationships between skill level and historical content. Students may show the ability to contextualize the events of September 11th but not D-Day. Using these formative assessments could also prevent teachers from reading too much into the results of summative assessments and standardized tests.

AP Historical Reasoning Skills and Disciplinary Practices

Like the broader trend in history education, the AP history courses and exams have decreased their emphasis on rote memorization and increased the weight of historical reasoning skills (Charap, 2015). Each of the four AP historical reasoning skills has three levels of proficiency described in the framework. The lowest level of proficiency simply requires students to be able to describe an example of causation or one of the other skills. The second level asks students to explain the historical phenomenon and the third level asks students to explain the significance of the historical phenomenon. Each question on the test, especially on the Long Essays, Document-Based Questions (DBQs), and short answer portions of the exam is supposed to assess one of the skills or practices. Although the College Board has responded to the shift toward more skills and less memorization, both the College Board and Charap (2015) have decided that multiple-choice questions still have a place—although it is a reduced one—in the assessment of historical thinking. Gabriel Reich (2015) at Virginia Commonwealth University believes that most multiple-choice exam results provide very little useful information for teachers but they continue to dominate the assessment landscape because they provide quick and quantifiable information that policymakers, district administrators, and politicians can use to make financial, curricular, and hiring decisions. Director of the College Board's Social Studies Content Advanced Placement Program Lawrence Charap (2015) claims that the AP history exams give educators good feedback about their students in relationship to their peers across the country and around the world because so many of them take the tests.

There is always an element of teaching to the test when there are high-stakes standardized assessments, but the changes to the structure of the AP courses and exam have had some of the intended effects on history instruction in American classrooms. Saul Straussman (2015), an

eleventh-grade APUSH teacher at Taylor Allderdice High School, reflected on the challenges of attempting to adjust to the new emphasis on skills instead of memorization of facts about U.S. history. Straussman chose to use a more thematic approach to teaching U.S. history that relied less on marching chronologically through a textbook while forcing students to memorize the seemingly endless list of possible names, dates, and events that could have appeared on the old test. Straussman's class now reflects the changes that the College Board hoped to see; students do more than learn facts, they learn to make connections, to identify and explore themes, to analyze and evaluate sources, and to construct their own arguments. They do the work of historians. Of course, they do not do it as well as professional historians, but high school calculus students do not do math at the same level as doctoral candidates in mathematics. Historians look for patterns and themes, and the new course with its seven themes encourages teachers and students to look for these patterns. The design allows teachers to continue to teach chronologically while pointing out connections to the different themes as they arise along the way (College Board, 2017a).

CHAPTER III

SOLUTION AND METHOD

Proposed Solution

The problem of practice is that our 8th grade history course does not have a system for tracking and assessing students historical reasoning and thinking skills. As a history department, we agree that the skills are more transferable than content from one grade level to the next. The skill that the 8th and 9th-grade history team agreed was most basic and essential to success in history courses at our school was document analysis. The proposed solution is to implement a document based lesson cycle two or three times per week for a full nine-week grading cycle. The document-based lesson cycle was developed based on the Model of Domain Learning (MDL), which assumes that expertise develops along a continuum (P.A. Alexander, 2003). The three phases of the MDL are Acclimation, Competence, and Proficiency. The acclimation phase is the part of the growth process when students learn the basics of a discipline or domain. In this phase, they learn basic vocabulary, concepts, and skills. The competence phase is characterized by a knowledge of the way information is organized within the discipline. Students learn how to use the domain-specific skills and heuristics. The proficiency phase, also known as the expertise phase, is reached when students begin to pose their own questions and make novel contributions to the field. One distinction between expertise and competence is that experts do not just know how to use the different skills and organizational tools in a particular discipline, but the different tools and organizational systems are integrated into the way the expert approaches and thinks about problems or issues within the domain (Alexander, 2003).

The Document-Based Lesson Cycle begins with accessing prior knowledge and activating situational interest. The next phase is direct instruction in which the teacher models expert thinking and analysis without input from the students. The application and integration phases allow students to practice using the skills the teacher modeled and to discuss in large or small groups what they have learned about the content and most importantly about the skills and tools used to analyze the documents (Stoel, 2015). The Document Based Lesson Cycle and the MDL Model both assume that students need explicit instruction and practice with the target skills.

There are eight sections of 8th Grade History at Independence Academy in the 2018-2019 school year. Four sections were in the control group and the other four sections were in the experimental group. Students in the control group received instruction that focused primarily on historical content. They had opportunities to analyze documents individually and in groups, but their guiding questions highlighted the acquisition of first-order knowledge or content instead of skills. For example, during our Radical Reconstruction lesson, outlined in Figure 1, the experimental group sourced and contextualized multiple primary sources written by Radical Republicans and Southern Democrats, while the control group answered content questions about sections of their eighth grade history textbooks and database articles. We added a few lessons since each unit was twice as long this year as it was last year. Instruction for the students in the experimental group focused on the acquisition of second-order knowledge that emphasized heuristics or cognitive structures and tools that historians use when analyzing historical documents and when thinking about historical issues.

Radical Reconstruction Lesson Outlines	
"Reading Like a Historian" Outline (Experimental Group)	Teacher-Developed Lesson (Control Group)
• **Focus Activity:** Photograph Analysis • (Students responded to prompts that encourage them to reflect on the origin and setting of the photograph. The emphasis was on the origin, nature, and reliability of the source.) • **Debrief** • **Interactive Mini-Lecture:** Review the transition from the Civil War to Reconstruction and the establishment of post-war political and economic landscape • **Document Analysis:** Answer sourcing questions about primary source documents from Andrew Johnson and Thaddeus Stevens • **Discussion Questions:** Students compare and contrast documents and discuss how the documents influenced their understanding and assessment of Radical Reconstruction.	• **Focus Activity:** Photograph Analysis • (Students and teacher analyzed photograph with an emphasis on acquiring content knowledge about Reconstruction. There was no discussion of the nature and reliability of the source.) • **Debrief** • **Interactive Mini-Lecture:** Review the transition from the Civil War to Reconstruction and the establishment of post-war political and economic landscape • **Textbook Reading:** Students read about the social and political conflicts between Southern Democrats and Radical Republicans in their 8th grade textbook and answered teacher created questions about them. • **Discussion Questions:** Students compare and contrast documents and discuss how the documents influenced their understanding and assessment of Radical Reconstruction.

Figure 1 Radical Reconstruction Lesson Outlines:
(Stanford History Education Group, 2019). Adapted.

Limitations

There are only about one hundred twenty students in the eighth grade at Independence Academy. This is a very small population. This makes it difficult to draw conclusions about the effect this type of intervention might have in other schools or with other students. The fact that the intervention only lasted six weeks is another potential limitation in the study. The benefits of limiting the length of the study were practical and ethical. First, it allowed time, if necessary, to alter the instructional method so that all students could receive direct instruction in document analysis and historical reasoning skills. Second, by limiting the length of the intervention, parents and students did not have to worry about some students being significantly better prepared for high school history classes simply because they were assigned to a specific section of the course. While there are practical reasons for limiting the length of the study and the

sample size is limited by the school's size, both of these factors decrease the overall reliability of the study (T.C. Urdan, 2017). However, since this is an internal study designed to improve instruction and assessment in a specific context the limited generalizability and the decreased statistical reliability of the results do not hinder the study's usefulness to our school community. The study is not designed to generate a new theory of history education or a program that can be applied in any secondary school. To the extent that the study gives insight into how to effectively teach and assess historical reasoning in eighth-grade classes at Independence Academy, the study will be considered a success.

Another major limitation of the study was that both teachers taught students in the control and experimental groups. Inevitably, the instructional strategies and approaches overlapped between the two groups because the same teachers planning and facilitating lessons for one group were also doing it for the other one. In an ideal setting, there would be enough participants to assign the control and experimental groups to different teachers and to control for the impact of individual teacher effectiveness and skill. However, action research often occurs at sites where a small number of teachers can participate in the study, and, for practical and ethical reasons, cannot vary instruction too dramatically between groups.

Justification of the Solution

The researcher chose to implement the Document-Based Lesson Cycle because document analysis is one of the most fundamental skills needed to succeed in history classes (Claravall, 2011). Document analysis and the ability to identify relevant evidence when writing historical arguments were the skills that the ninth grade teachers at Independence Academy said they work on the most when helping freshmen transition to their upper school history classes. The Document-Based Lesson Cycle aligns well with the MDL because the cycle allows time for the

teacher to model expert thinking and provides time for students to work in small and large groups as apprentice historians (G.L. Stoel, 2015).

Proposed Research Paradigm

This study was primarily a quantitative study that sought to compare the changes in performance on assessments of historical thinking between the students in the implicit and explicit conditions. The researcher wanted to determine if one method of instruction yielded better results than the other. Some of the data was qualitative because students answered questions about historical documents. Their answers to the questions had to be interpreted and scored. The answers themselves are qualitative in nature, but the scoring was numeric, as the answers will be graded using rubrics that assign different point values to different skill levels.

Data Collection Methods

Students in both the implicit and explicit condition took a four-part pretest and a posttest. The first section of the pretest and corresponding section of the posttest measured recall of historical facts. We used this content-focused assessment to identify any differences in background knowledge between the different classes and to establish a baseline score for each student. In order to see if there was a difference between the amount of factual knowledge gained during the intervention period, we gave students the same twenty content questions they answered on the pretest. The questions required the students to write direct answers to close-ended questions. Three other question types helped us measure students' ability to analyze and use evidence from historical documents.

Weighted Multiple Choice Questions (WMCs)

The first skills assessment consisted of Weighted Multiple-Choice (WMC) questions. Each of these questions targeted a specific historical thinking skill. Instead of having one correct

answer, a WMC question has at least three answers that are historically defensible. However, the different "correct" answers represent varying skill levels (VanSledright, 2015). Presumably, students with more advanced historical thinking skills selected the more heavily weighted answer choices more often than their peers. The most advanced answer choices received four points, the next level of competence earned three or two points, and the novice level answers received one point. Totally incorrect or indefensible answer choices earned zero points. Figure 2 is a WMC taken from the UMBCs Center for History Education's online bank of history assessment items. This particular question requires students to analyze two primary documents about Westward Expansion and attempts to measure their ability to read critically and corroborate. The number in parentheses indicates how many points students would earn by selecting that particular answer. Students answered ten to fifteen WMC questions before and after the intervention.

Source 1: Thomas Jefferson to John Breckinridge [Senator], 12 August 1803

This treaty must of course be laid before both Houses, because both have important functions to exercise [apply] respecting it. They, I presume [believe], will see their duty to their country in ratifying [approving] & paying for it, so as to secure a good which would otherwise probably be never again in their power. But I suppose they must then appeal [ask] to the nation for an additional article to the Constitution, approving & confirming an act which the nation had not previously authorized. The Constitution has made no provision [prearrangement] for our holding foreign territory, still less for incorporating [admitting] foreign nations into our Union.

Source: Thomas Jefferson Papers. Library of Congress. Washington, D.C. Accessed 5/18/14. http://memory.loc.gov/cgi-bin/ampage?collId=mtj1&fileName=mtj1page028.db&recNum=1042

Source 2: Thomas Jefferson to Wilson Cary Nicholas [Senator], 7 September 1803

. . . Whatever Congress shall think it necessary to do should be done with as little debate as possible, and particularly so far as respects the constitutional difficulty. I am aware of the force of the observations you make on the power given by the Constitution to Congress to admit new states into the Union, without restraining [limiting] the subject to the territory then constituting [making up] the U.S. But when I consider that the limits of the U.S. are precisely fixed by the treaty of 1783, that the Constitution expressly declares itself to be made for the U.S., I cannot help believing the intention was to permit Congress to admit into the Union new states which should be formed out of the territory for which, and under whose authority alone, they were then acting.

Source: Jefferson, Thomas to Wilson Cary Nicholas, January 26, 1802. Founders Online, National Archives and Records Administration, Washington, D.C. Accessed 5/18/14. http://founders.archives.gov/documents/Jefferson/01-36-02-0280

Question: Based on these two sources, what did Thomas Jefferson believe about the constitutionality of the Louisiana Purchase?

A. He believed that the Constitution was flexible enough to allow for situations not anticipated by the founders, like the Louisiana Purchase. (0)
B. He believed that the United States must return Louisiana to France since the Constitution intended to limit the creation of new states to those territories existing in 1783. (2)
C. He believed that the purchase of Louisiana was unconstitutional but that the future benefits it presented required the amendment of the Constitution to allow it. (4)
D. He believed that the Constitution left it entirely up to Congress to decide whether to permit the purchase of Louisiana. (1)

Figure 2 Westward Expansion WMC: (UMBC Center for History Education, 2018). Reprinted.

History Assessments of Thinking (HATs)

The second portion of the skills tests required students to complete twelve History Assessments of Thinking (HATs) (Wineburg, Smith, and Breakstone, 2018). These short assessments were designed to target specific historical thinking skills. Like WMCs, HATs usually ask students an analytical question about an image or excerpt from a primary or secondary source. Unlike WMCs, HATs are open-ended and do not provide students with answer choices (Stanford History Education Group, 2018). The answers were graded using a rubric that measures the proficiency level of the students' responses. For this study, used the 3-point rubrics provided on the Stanford History Education Group website and others developed by the researcher. HATs target skills like sourcing, corroboration, use of evidence, and contextualization. One HAT, shown in Figure 3 below, developed by Smith, Breakstone, and Wineburg (2018) tests students' sourcing skills by asking them to evaluate the reliability of a 1932 painting of the first Thanksgiving.

Directions: Use the painting to answer the question below.

Title: "The First Thanksgiving 1621"
By: J.L.G. Ferris
Date painted: 1932

Question: The painting, "The First Thanksgiving 1621," helps historians understand the relationship between the Wampanoag Indians and the Pilgrim settlers in 1621.

Do you agree or disagree? (Circle one).

Briefly support your answer:

Figure 3 "The First Thanksgiving" HAT: (Smith, Breakstone, and Wineburg, 2018). Reprinted.

Document-Based Questions (DBQs)

The final portion of the skills assessments took the form of Document-Based Questions. Document-Based Questions ask students to write an analytical essay using evidence from several primary sources which the teacher or test creator provides. DBQs are not only used in many history classes in secondary schools, but they are also a major component of the A.P. History Exams (Charap, 2015). Unlike the other two skills portions of the pretest and posttest, the DBQs required students to combine a variety of historical reasoning skills instead of focusing on one particular skill. I originally intended to grade the DBQs using rubrics provided by the College Board but the ARCH rubric, included in Figure 4, is more closely aligned with the target skills of this intervention. The ARCH rubric also provides short descriptions of each level of proficiency, which made it easier to be more consistent in the grading process (UMBC Center for History

Education, 2018). A limitation to using this rubric was that it was not designed specifically for grading DBQs and our students were unfamiliar with it.

Historical Thinking Skills Scoring Rubric – Secondary

	Close Reading Strategies		Strategies/Procedural Concepts		Procedural Concepts	
Criteria	**Sourcing**	**Critical Reading**	**Corroboration**	**Contextualizing**	**Claim**	**Evidence**
4	***Identification:*** Fully understands the meaning and content of sources. ***Attribution:*** Cites all authors and all original dates of primary and secondary sources. ***Perspective:*** Evaluates the reliability sources based on the author's perspective and when and why they were produced.	▪ Questions the author's thesis and determines viewpoint and evidence to evaluate claims, highlighting what the author leaves out. ▪ Cites accurate examples of how the author uses persuasive language and specific words and phrases to influence the reader. ▪ Seeks answers to questions left unanswered in the source to formulate an interpretation.	Constructs an interpretation of events using information and perspectives in multiple sources. Identifies consistencies and inconsistencies among various accounts.	Applies prior and new knowledge to determine the historical setting of sources. Uses that setting to interpret the sources within the historical context as opposed to a present-day mindset.	Formulates a plausible interpretation, argument, or claim based on the evaluation of evidence found in a variety of primary and secondary sources.	Justifies claims using appropriate direct evidence from a variety of reliable sources.
3	***Identification:*** Mostly understands the meaning and content of sources. ***Attribution:*** Cites most authors and most original dates of primary and secondary sources. ***Perspective:*** Examines the reliability of sources based on the author's perspective and when and why they were produced.	▪ Analyzes the author's thesis, determines the viewpoint and evidence to evaluate the claims; may highlight what the author leaves out. ▪ Cites examples of how the author uses persuasive language and specific words and phrases to influence the reader. ▪ Notes that the author has left some questions unanswered.	Explains similarities and differences by comparing information and perspectives in multiple sources.	Applies prior and new knowledge to determine the historical setting of the sources. May attempt an interpretation of some sources with a present-day mindset or with a limited application to the historical context.	Generates a reasonable interpretation, argument, or claim based on an evaluation of the evidence found in selected primary and secondary sources.	Justifies claims using some appropriate direct evidence from a variety of reliable sources.
2	***Identification:*** Understands the meaning and content of sources with appropriate scaffolding and support. ***Attribution:*** Cites some authors and some original dates of primary and secondary sources. ***Perspective:*** Attempts to evaluate the reliability of sources.	▪ States the author's claims and evidence presented to prove those claims. ▪ Determines the author's viewpoint. ▪ Notes how language is used to persuade.	Identifies similarities and differences in information in multiple sources.	Attempts to determine the historical setting of sources without fully understanding the historical context.	States an interpretation, argument, or claim that may or may not based on evidence found in selected primary and secondary sources.	Justifies claims using generalizations or limited appropriate direct evidence.
1	***Identification:*** Attempts to understand the meaning and content of sources with the appropriate scaffolding and support. ***Attribution:*** Cites few authors and few original dates of primary and secondary sources. ***Perspective:*** Does not adequately examine reliability.	Attempts to identify the author's claims, viewpoint, or evidence.	Demonstrates little to no attempt to examine sources for corroborating or conflicting evidence.	Demonstrates no attempt to understand the historical setting of sources.	Does not state an original claim, argument, or interpretation.	Does not justify or support claims using appropriate direct evidence.

UMBC Center for History Education, 2013. Adapted from the work of the Stanford History Education Group * and Bruce VanSledright, *Assessing Historical Thinking and Understanding: Innovative Ideas for New Standards,* (New York: Routledge, 2014).

Figure 4 ARCH Secondary Historical Thinking Skills Rubric: (UMBC Center for History Education, 2018). Reprinted.

Justification of Use of Instruments

Although most historians and history educators consider rote memorization of historical facts to be the lowest level of historical thinking, they still recognize the importance of facts and evidence to the work of historians (Reich, 2015). Historical reasoning is not done in a vacuum void of specific historical circumstances, so historical thinking skills can and should be taught and assessed as a part of the study of specific historical events, movements, and issues (Smith,

2017). For this reason, the students in this study were assessed on their first order historical knowledge or their knowledge of specific historical facts. Another reason why this study included content-focused assessments was to see if emphasizing historical reasoning skills led to a decline in the amount of factual information students learned and retained during the intervention.

The WMCs were designed to compensate for some of the weaknesses of traditional multiple-choice questions, while still allowing for easy grading. VanSledright (2015), a leader in the field of assessing historical thinking, promotes the use of WMCs because they are intended to measure historical thinking instead of just first-order knowledge of facts. The answer choices represent different levels of historical thinking. VanSledright does not spend much time focusing on the fact that students could just as easily guess on a WMC as they could on a traditional multiple choice question. Though he does not address this possibility, his writing suggests that students will most consistently select answer choices that reflect their skill level. This question type could decrease in effectiveness as students answer and receive feedback on more of them. Over time, students could just learn to pick out the language that seems most complex or to look for certain keywords that they associate with higher levels of historical thinking. This would be different than being able to produce a written response that reflects a higher skill level. To avoid the problems that could emerge from overuse of WMCs students only answered these types of questions on the posttest and on the pretest, which they did not receive a copy of until after the intervention and final assessments are completed. To check for the validity of the WMCs, I originally planned for students to write one or two sentences justifying each of their WMC answer choices. However, the skills portion of the test took much more time than expected and students were not required to justify their answers on this section.

HATs are designed to decrease the chances that students can guess the correct answer. In fact, HATs do not ask students to recall factually correct information. Students are given a primary source to analyze. The question that accompanies the source asks students questions that allow them to demonstrate their historical reasoning skills (Wineburg, Smith, and Breakstone, 2018). Students might be asked to evaluate the reliability of a source based on the background of its author, the date of its production, or the historical context in which it was produced. Since no answer choices were given students generated their own responses to the questions and explained how they reached their conclusions. Sam Wineburg and the Stanford History Education Group have developed a bank of HATs that are available for free on their website (SHEG, 2018). Some of the HATs used in this study came directly from the SHEG website and the researcher used other questions from the site as models for developing the rest of the HATs.

Document-Based Questions have their problems, but they are a useful teaching and assessment tool in secondary history classes. Bruce VanSledright (2015) criticizes the fact that DBQs attempt to measure multiple skills in one exercise. He suggests that teachers, students, and researchers gain little information about students' progression on any one skill because so many of them are assessed simultaneously. It is difficult to isolate any one skill in an essay that requires so many steps and complex cognitive processes to complete. DBQs are also time-consuming for most teachers to grade. They are also difficult to create for teachers that choose to build their DBQs from scratch. The question itself may be simple to write, but the gathering of the primary documents that would be most appropriate for the task can be a particularly arduous task. While they can be time-consuming to create and grade and, as VanSledright (2015) notes, they measure a wide range of skills, DBQs are some of the most authentic tasks

novice historians can complete in a limited time period. Like professional historians, students who are completing DBQs analyze a variety of sources in an attempt to answer a historical question. Unlike professional historians, students do not generate their own questions or hunt down their own sources. Professional historians are not under the same time constraints as students and teachers in secondary classrooms. The DBQ is not a purely authentic assessment of historical reasoning skills, but it is about as close as one can get in the confines of a classroom testing environment. The fact that many Independence Academy students will take at least one AP History Exam is another reason to evaluate the impact of instructional methods on student performance on DBQs.

Data Analysis Strategy

The main data points for this study will be student performance scores on their pretests and posttests. The independent variable in this study was the instructional method used in the different groups. Some received implicit instruction in document analysis and historical thinking, while others will receive explicit instruction in these skills. The dependent variables in this study were the performance on the measures of content knowledge and skills. The students received a score of 0-20 on the content assessment. They received one point for each correct answer. Students received a score of 1-4 on each WMC and a score of 1-4 on each HAT. The researcher calculated the average score for each student's performance on each question type. The students wrote one essay response to a different DBQ before and after the intervention.

The researcher conducted an ANCOVA to see what impact there was, if any, of the instructional intervention on student performance on content questions, WMCs, HATs and DBQs (Urdan, 2017). Each of these question types represented a different dependent variable. The use of ANCOVA was selected because they are helpful when trying to determine the

amount of variance between dependent variables such as test scores that share an independent variable, which in this study would be the assigned group (Urdan, 2017). I hypothesized that there would be little to no difference in performance between the control and experimental groups in terms of their performance on the measure of content knowledge. I further hypothesized that both groups would show growth in content knowledge. Finally, I also hypothesized that students in the explicit condition would outperform the other students on the three measures of historical thinking skills.

Intervention Timeline

No student names were shared with anyone other than the middle school administration and none were included in this record of study. Students took the content and skills pretest the week of January 28, 2019. They also began their six-week study of American Identity. Each week they studied a different social movement or issue that relates to how Americans have defined themselves. Independence Academy's middle school uses a modified block schedule. On Mondays and Fridays students attended eight 37-minute classes and on Tuesday, Wednesday, and Thursday they attended six 60-minute classes. They attended each core academic class four times each week. The researcher and the other eighth-grade history teacher planned to implement the Document-Based Lesson Cycle on each block day for the sections of the 8th-grade course that will be in the explicit condition. After using two instructional days for the Reconstruction Quiz and the American Identity Test, we were left with fifteen actual lessons. Eight of the lessons (or at least part of them) for the intervention group were taken directly from the Reading Like a Historian online curriculum. This means that there was at least a 53% fidelity rate to the DBL Model. At least two or three of the other lessons were designed by the

teachers to fit the same model or were extensions of the RLH Lessons. The dates for the various units and assessments are outlined below on Figure 5.

Dates	Activity
January 28-30	Pretests and Introduction to New Theme
January 30-February 8	Unit 1: Reconstruction
February 11-21	Unit 2: Gilded Age/Populism
February 22-27	Unit 3: Progressive Era
February 28-March 4	Unit 4: Great Depression
March 5-7	Posttests

Figure 5 Intervention and Assessment Timeline

Reliability and Validity Concerns

Complex cognitive processes, like historical reasoning, are notoriously hard to teach and even more difficult to assess (VanSledright, 2014). One of the biggest challenges in assessing historical thinking is that students frequently use skills that are not specific to any particular domain to answer history assessment test items. History educators have responded by developing assessment tools intended to elicit specific cognitive processes. To test the validity of these assessment tools history educators, like Wineburg and VanSledright, have turned to Thinking Aloud Protocols (TAPs), which are interviews that occur during or after an assessment in order to give the researchers insight into the thinking of the test takers. The researchers ask the test takers to talk about what they are thinking as they read, process, and attempt to answer test questions (Ercikan & Seixas, 2015). These interviews sometimes reveal that test items are not valid because so many of the students who responded correctly used nonhistorical tools to select or construct their answers. G. A. Reich (2015) found in a study that evaluated history

assessments that the traditional multiple choice question is the least valid form assessment. HATs and WMCs decrease the chances of students guessing and give the teacher insight into the rationale and historical reasoning skills they used to answer the questions. When they use skills that are not history specific it becomes evident because they either tend to select answer choices that have lower point values or they construct answers that rate lower on historical thinking skills rubrics.

The DBQ may not be the most valid assessment of any particular historical thinking skill because it is such a large assignment with so many components. However, DBQs do force students to engage in activities that resemble the work of actual historians. VanSledright (2014) argued that DBQs are not very helpful because they require the analysis of multiple documents but most students need more help learning how to first analyze a single document. This may be true, but the purpose of including DBQs in this study is to see if practice analyzing individual documents over the course of an eight-week intervention improves eighth grade students' abilities to analyze and use multiple documents to construct and support their own historically defensible arguments.

CHAPTER IV

ANALYSIS AND RESULTS

As a part of this study, students took four pretests and four posttests. The differences between the two groups in terms of gains and overall performance were not statistically significant. Only 120 students were tested and many students' scores were omitted from the study because they missed or failed to finish one or more of the assessments. Ninety students finished all four portions of the pretest and posttest, and those ninety students' scores are the ones that are represented in this study. The intended purpose of the study was not to generalize the findings, but to compare the effectiveness of two particular teaching approaches in a specific context and to identify any relationships between instruction and student development of historical reasoning skills. There were some anecdotal observations and a couple of statistically significant findings on specific portions of the posttest that indicate a positive relationship between the use of DBLs and improved student reasoning skills.

Data Analysis Procedures

I graded each of the four pretests and posttests and then input the points earned on each content, WMC, and HAT question and on each category of the DBQ rubric onto a spreadsheet that I later imported to SPSS. The SPSS software allowed me to run independent samples t-test for the total number of points earned on each assessment and the pre/posttest gains and means. These *t* tests helped me to check for differences in the performance and progress between the treatment and control groups. After running these initial statistics, I also used SPSS to run Analyses of Covariance that compared the differences in performance and progress on the posttests while controlling for performance on the pretest. Finally, I identified the Pearson

correlation coefficients for the posttest results and portions of the standardized test that Independence Academy Middle School students take at the end of each year.

Results of Research

Research Questions:

1. How effective was the Document Based Lesson Cycle at improving eighth grade Independence Academy students' ability to analyze primary and secondary documents on history assessments?
2. What is the relationship between explicit instruction on document analysis in eighth-grade history classes and achievement on assessments of document analysis skills?

To help answer these questions, the students were first given a four-part pretest to determine if there were any differences in the skill-levels or content knowledge of the treatment and control groups that might make it more difficult to make inferences about the impact of the intervention. After the intervention, I compared the performance of the two groups on the four sections of the posttest and the gains they made over the course of the intervention.

Results of content knowledge assessments

The content knowledge pretest and posttest were identical. They both contained twenty recall questions about Reconstruction, the Gilded Age, Populism, the Progressive Era, the Roaring 20s, and the Great Depression. Students received one point for each correct answer and zero points for incorrect answers. Students received half of a point if they correctly answered part of a multi-part question. The mean score on the content pretest for the experimental group was 5.0 out of 20, while the control group earned 4.650 out of 20 points. When tested on the

exact same questions six weeks later, the control group's mean score was 16.128 and the experimental group's mean score was almost identical at 16.120 out of 20.

The control group began with lower scores and finished with higher scores, but Table 1 reveals below that there were no statistically significant differences in gains or overall performance between the two groups before or after the intervention. However, both groups improved by an average of 11.279 points, which indicates that both groups increased their content knowledge. A paired samples t test, summarized in Table 2 below, produced a significant *t* value (t (89) = 43.544, p <1). Although, the two research questions did not specifically address the acquisition of content knowledge, it was important to observe whether or not additional time spent on skill development contributed to a decline in the amount of content learned when compared with the control group. The content knowledge analysis led me to conclude that neither group entered the intervention period with a stronger knowledge base than the other group and that the skills-focused intervention did not prevent students in the experimental group from obtaining and retaining content knowledge at the same rate as their peers.

	N	Pretest Mean	Posttest Mean	Gains Mean	Standard Dev For Gains	Significance (2-tailed) for gains
Control	40	4.650	16.128	11.478	2.347	.496
Treatment	50	5.000	16.120	11.120	2.555	

Table 1 Content Test Results

	Mean	St. Dev.	St. Error Mean	t	Degrees of Freedom	Sig. (2-tailed)
Posttest-Pretest	11.2789	2.4573	.2590	43.544	89	.000

Table 2 Paired (Dependent) Samples *t* test for Significance of Difference Between Content Pretest and Posttest Scores

Results of Weighted Multiple Choice questions

The second assessment that was used to determine the effectiveness of the intervention and to establish if there was a relationship between student performance and instructional method was the Weighted Multiple Choice Question, each of which attempted to target a specific historical reasoning skill. The best answers received 4 points, the second best answer received 2 points and totally incorrect answers received 1 point. Table 4 shows that the control group's mean changed by 0.120 points from 2.763 to 2.883, and the experimental group's pretest and posttest means were 2.866 and 2.888 respectively. Neither the amount of growth within each group nor between groups was statistically significant so there was essentially no measurable change over the course of the intervention. Even when looking at individual questions or questions organized by skills they assessed (contextualization, corroboration, critical reading, and sourcing), the results for each group were very similar. As far as improving the ability of students to answer skills-based WMCs is concerned, this intervention was ineffective. In terms of the second research question, the results suggest that there was no relationship between the instructional method and students' abilities to answer WMCs that test sourcing, corroboration, contextualization, and critical reading skills.

	N	Pretest Mean	Posttest Mean	Gains Mean	St. Dev. For Gains	P value Sig. (2-tailed)
Control	40	2.763	2.883	0.120	0.447	0.395
Treatment	50	2.866	2.888	0.022	0.590	

Table 3 WMC Test Results (Independent Samples t test Comparing Gains)

Results of History Assessments of Thinking

The students answered twelve HATs on the pretest and thirteen on the posttest. The performance between the two groups was very similar on both tests. Student scores for this assessment were calculated by taking the average number of points earned out of three on each HAT. The control group mean changed from 2.126 to 2.444 and the experimental group improved from 2.175 to 2.474. This represented a 14.8% increase for the control group and a 12.6% increase for the experimental group. As Table 4 highlights, the difference in gains between the two groups was not statistically significant. When running a paired samples t-test on the mean gains for both groups, there was a statistically significant *t* value (t (86) = 7.568, p > .001).

	N	Pretest Mean	Posttest Mean	Gains Mean	St. Dev. For Gains	Significance (2-tailed) for Gains between Group
Control	40	2.126	2.444	0.315	0.342	0.601
Treatment	50	2.175	2.474	0.274	0.377	

Table 4 HAT Test Results (Independent Samples t test Comparing Gains)

After looking at these results, I have to conclude that both groups improved over the course of the intervention, but I must also conclude that neither one of the instructional approaches proved to be more effective than the other. Since this study did not record the amount of progress made in a normal six-week period, it is difficult to know if the students would have made the same amount of progress even if we had made no changes to the curriculum or instructional approach. Pearson correlation coefficients revealed that the correlation between pretest and posttest HATs scores were strong with a calculated *r* value of .500 and a *p* value that was less than .01. However, the coefficient of determination ($r^2 = .250$) indicates that much of the variance on the posttest cannot be explained simply by observing individuals' performance on the pretest. It is also difficult to know if the students actually made progress or if the posttest was simply an easier assessment than the pretest. The two tests tested the same skills, but they used different content and different questions.

Results of Document-Based Questions

Unlike the other three assessments, the DBQ did provide some evidence that the use of the DBL Cycle had a positive impact on at least one of the students' historical reasoning abilities. Only one DBQ was asked at the end of each test. The pretest DBQ asked students to explain how and why U.S. foreign policy evolved from the beginning of the 20th century to the present. Students were encouraged to use the photographs, documents, charts, and cartoons included in the HATs section of the pretest. They were also allowed to use evidence from other sources that we used in class during our study of War and Diplomacy. Each quarter of our course had a different theme. The pretest was given right after we ended the War and Diplomacy theme with the study of the War on Terror. The same format was used for the posttest HATs and DBQ, but the sources were related to the topics and time periods studied in the intervention

period. The DBQ for the posttest asked students to explain how and why the relationship between the U.S. government and its citizens evolved between 1865 and 1935. The HATs that preceded the DBQ included textbook excerpts, quotes from speeches, population charts, political cartoons, and photographs from the Reconstruction Era through the Great Depression.

The DBQ scores were calculated by taking the 4-point scale score for each of the categories on the ARCH rubric and then dividing by six since there were six skills measured by the rubric. The mean DBQ for the pretests was 2.346 for the control group and 2.383 for the experimental group. On the posttest test, the treatment and control group improved to 2.629 and 2.833 respectively. The calculated p value of .487 on Table 5 below indicates that the difference between the gains of the two groups was not statistically significant. However, the same is not true when looking at the first skill measured by the rubric, the ability to make claims using historical evidence from a variety of different sources.

	N	Pretest Mean	Posttest Mean	Gains Mean	St. Dev. For Gains	Significance (2-tailed)
Control	40	2.346	2.629	.2833	.79636	.487
Treatment	50	2.383	2.833	.3949	.68616	

Table 5 DBQ Test Results (Independent *Samples t* test Comparing Gains)

The results of the DBQ provide evidence that there was a connection between the type of classroom instruction students received and their ability to make historical claims. Prior to the intervention, the control group had a 2.80 mean score on the Claim category of the ARCH rubric (p. 40), while the experimental group had a 2.76 mean score. In order to receive full credit (4

points) on this part of the rubric students had to "formulate a plausible interpretation, argument, or claim based on the evaluation of evidence found in a variety of primary and secondary sources" (UMBC Center for History Education, 2018). After the intervention, the control group's mean score was 3.43 and the experimental group's score was 3.74. When running an ANCOVA for the posttest scores that controlled for pretest scores, the difference between the two groups had a calculated value of .014, well below the .05 needed to indicate statistical significance, and a partial eta of .070, which suggests that placement in any given group accounted for about 7% of a student's score on this skill.

A challenge to assessing historical reasoning is that many history tests are actually measures of reading ability or other skills that are not specific to history as a discipline (VanSledright, 2015). This study seems to give some credence to that idea. The instructional method was not a strong predictor of student performance on the posttest. However, when looking at the Pearson correlation coefficients for our eighth graders' Educational Records Bureau (ERB) standardized exam scores there were some moderate and strong statistically significant correlations between the ERB scores and the posttest scores. Timothy Urdan, author of *Statistics in Plain English* (2017), explains that Pearson product-moment correlation coefficients measure the direction (positive or negative) and the strength of relationships between continuous variables, such as test scores. Table 6 below shows the *r* values for the Pearson correlation coefficients for ERB test scores and the posttest means for Independence Academy eighth graders. There was a strong or moderate positive correlation between ERB scores and posttest performance. The only exception to this was the relationship between the ERB Writing Mechanics score and the DBQ score; for these two measures, the correlation was positive but weak. Of the correlations, this was the only one that was not statistically significant at the 0.05

confidence level. These numbers reveal that students who performed better on the verbal sections of the ERB also performed better on the different assessments of historical thinking given in this study.

	ERB Verbal Reasoning	ERB Vocabulary	ERB Reading Comprehension	ERB Writing Mechanics	Assigned Group
Content	.446** Moderate Correlation	.593** Strong Correlation	.541** Strong Correlation	.476** Moderate Correlation	-.002 Weak Correlation
WMC	.406** Moderate Correlation	.449** Moderate Correlation	.389** Moderate Correlation	.328** Moderate Correlation	.005 Weak Correlation
HAT	.517** Strong Correlation	.365** Moderate Correlation	.443** Moderate Correlation	.518** Strong Correlation	.048 Weak Correlation
DBQ	.213* Moderate Correlation	.248* Moderate Correlation	.241* Moderate Correlation	.134 Weak Correlation	.154 Weak Correlation

** Correlation is significant at the .01 level
*Correlation is significant at the .05 level
Table 6 Pearson Correlation Coefficients for ERB Scores and Posttest Means

There was no statistically significant correlation, using Point Biserial Correlation (a special form of Pearson correlation), between the group that students were assigned to and their scores on the different sections of the posttest. This supports the claim that the assessments are, at least in part, measures of reading and test taking skills. Even though the correlations suggest a relationship between standardized test scores and posttest performance, other factors such as instructional methods, gender, teacher, class size, content-area interest, and historical thinking skill level prior to the intervention, could have contributed to student outcomes.

After finding the Pearson product-moment correlation coefficients for the different sections of the pretest and posttest, I was reminded of the studies by Buckingham (1917) and by Bell and McCollum (1919), which both argued that historical thinking skills assessments

produce results that are similar to measures of historical content knowledge. The results in my study did not fully support their findings. There were moderate positive correlations between the content knowledge pretest results and the other three sections, which are shown in Table 7. This makes sense because it is easier for students to answer questions of authorship, context, meaning, or connection to other sources when they have background knowledge about the events. One reason why the correlations might not have been stronger is because Buckingham, Bell, and McCollum asked different kinds of questions and early in the 20th century viewed historical thinking differently than more recent scholars and test makers. Their recall questions may have been similar to the content questions we asked, but their higher level thinking questions may have focused more on analyzing content and drawing conclusions from historical texts. Many of the questions we asked required students to evaluate the usefulness or reliability of the sources. This difference in emphasis could have weakened the correlation between content knowledge and performance on skills assessments.

	WMCs	HATs	DBQs
Content Pretest	.299** Moderate Correlation	.443** Strong Correlation	.284** Strong Correlation
Content Postest	.364** Moderate Correlation	.405** Moderate Correlation	.208 Moderate Correlation

** Correlation is significant at the .01 level

Table 7 Pearson Correlation Coefficients for Content and Skills Assessments

Summary of Pretest and Posttest Data

None of the assessments revealed statistically significant differences between the control and experimental groups in growth or overall performance. However, the experimental group's

performance and growth on the Claim category of the DBQ and on one of the Photo Analysis questions on the HATs suggest that the instructional approach or something about the classroom experience impacted how the students approached these assessments. As the Literature Review explained, effective assessments of historical reasoning are difficult to create and performance on them can be influenced by many other factors since historical thinking is such a complex cognitive process. It is totally possible that the impact of the instructional approach was diluted by everything from time of day, reading ability, fatigue, and anxiety to writing skills, interest in the topics, and the amount of effort given on these assessments that students knew would not have a major effect on their grades in the course.

Interaction between the Research and the Context

How did the context impact the results?

Several characteristics and practices at Independence Academy may have had an impact on the results. The Middle School has, for the past two school years, used a modified block schedule, shown below in Figure 3, with 37-minute classes on Mondays and Fridays and 60-minute classes the rest of the week. Students report to each core academic class four times a week. This means that in addition to having history class at different times of day, students had different lessons on different days of the week. In some cases, different classes had different lessons on the same day of the week. Because of the rotation in the schedule, classes met at different times on different days of the week. This added a complicating factor to the scheduling, planning, and delivering of lessons and assessments. This schedule also meant that the pre- and posttest were administered to different classes at different times and on different days.

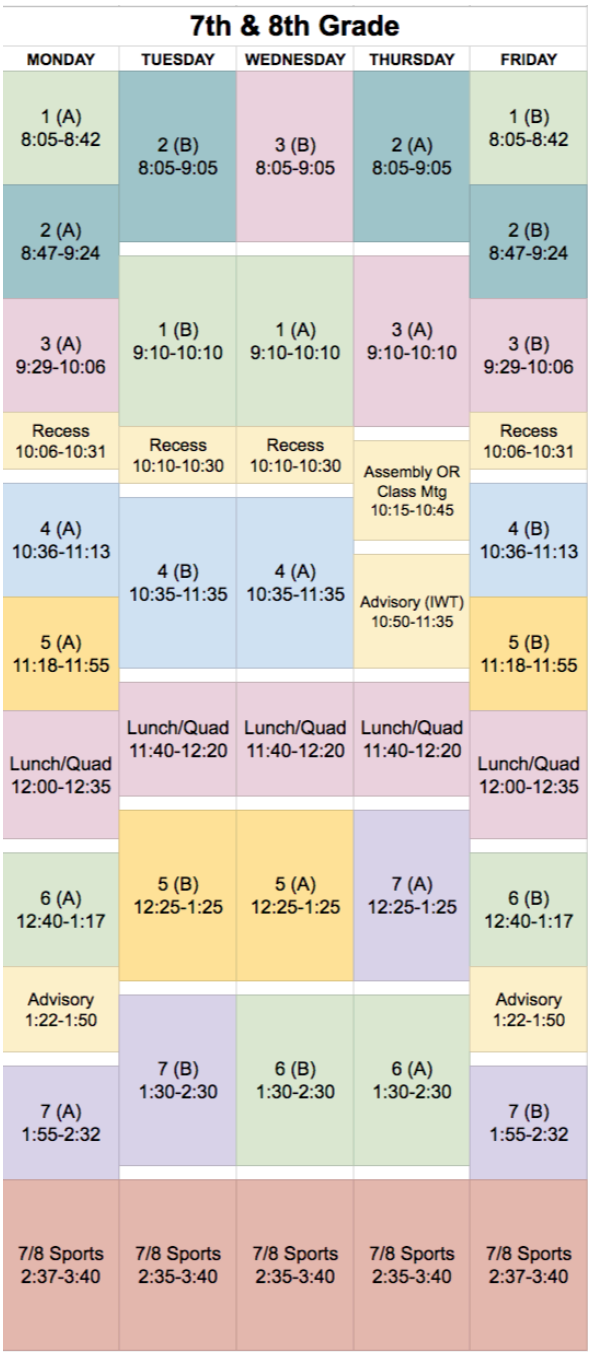


Figure 6 Independence Academy Middle School Modified Block Schedule.

During the school year, teachers at Independence Academy encourage students not to discuss assessment questions and answers with their peers before all groups have had the opportunity to take the assessment. However, we are certain that some students could not resist

the urge to share their thoughts about this series of assessments. Some students may have shared some of the simple content questions with their peers and those students could have very easily looked up the answers or asked someone who was confident about the correct answer. It would have been more difficult to look up the correct answer to the other sections of the tests, but the integrity of the results may have been compromised by this unofficial and unsanctioned collaboration among students.

The two eighth grade history teachers decided not to count the content pretest as a grade since it was over previously uncovered material and even though the content of the WMC, HAT, and DBQ sections had been covered in the previous unit, the teachers decided to count these sections as regular classwork grades. Some students admitted to other students and to the teachers that they did not give their best effort on the pretest or posttest because they knew that their performance would not dramatically impact their grades. The researcher cannot be sure that the results for individual students or the group represent their best effort.

Another challenge that arose in administering and grading the assessments was the high number of student absences and tardies. The posttest was given the week before Spring Break so student engagement, interest, and focus were not at their peak and many students were absent. Of the 120 students enrolled in the course, only about 90 of them had complete data for all four pre and posttests. Some students finished all sections of the pretest but not all sections of the posttest and vice versa. The reasons for not finishing vary from student to student. Some openly expressed frustration and anger about having to spend two, and in some cases two and a half, class periods taking assessments that would not greatly affect their grade in the course. Some students were sick or out of town on the day of the assessment. Others had to miss class because of other meetings, rehearsals, or athletic activities on campus. About 30 of the students have

academic accommodations that allow them to have extended time on assessments. However, because this assessment was not counted as a test grade, several students did not take advantage of their right to extended time. Only the results for the 90 students who completed the pretests and posttests were not counted in this study.

The Interim Head of School, the middle school administrators, and the middle school history department all supported the implementation of this study. Each year, teachers at Independence Academy develop two goals for professional growth, and their administrative evaluator assigns them another goal. One of the goals that the middle school head gave me for the 2016-2017 school year was to create a list of the skills that were taught in our 8th grade history course. This research is linked, although a few years late, to that goal.

Independence Academy is a pretty close-knit community with high levels of parental involvement. Even in the absence of state-mandated standardized exams, parents and students are interested in grades, high school course recommendations, and course curricula. The expectation in the middle school is that students in different sections of the same course will have comparable experiences. The two history teachers chose to post the same assignment and homework sheets for the control and experimental groups during the intervention period and handouts from both groups were made available to all students on the teachers' websites. Without the need to make sure that the students had similar experiences, there would have been even greater pedagogical differences between the two groups. In most lessons, the experimental group looked at a variety of primary documents and the teacher used guided questions from the RLH lessons. The control group usually analyzed two or three secondary sources and teachers asked content-focused questions of the text and did not model historical reasoning strategies. The similar structure of the lessons may have contributed to the similarity in outcomes. Even

without the teacher modeling the historical reasoning skills, students in the control still had opportunities to analyze, compare, and contrast various historical accounts. Since there were students with varying levels of historical reasoning skills in each class, students had chances to see others model higher levels of analysis even though the teacher did not conduct think alouds for them. Ideally, the structure and sources used in the lessons would have been totally different in order to create more contrast between the instructional experience of the groups. At the same time, thanks to the similarity in structure, the main differences between the experiences of the two groups were the types of questions asked of the students, the types of sources analyzed in class, and the amount of modeling done by the teacher. It is not possible to determine for this particular study how those factors interacted with one another or how they might have influenced the results

A factor that was not really considered before beginning the study was the amount of time it took to actually grade the assessments. The grading of the content portion of the pretest was pretty simple and was finished within a day or two after administering the test. However, the rest of the pretests were not fully graded until Spring Break, after the intervention was over. This is obviously not ideal because the teachers were not able to use the results of the pretest to guide their instruction because they did not have the results. Students were not able to use their pretest results to guide their practice, study, and preparation for the intervention lessons and posttest. Finally, the researcher was the only one who graded the different sections of the tests. While two other teachers gave feedback on the weighting and answers on the WMCs, the bulk of the evaluation of student work was done by one person. There was also a conflict of interest because, as the researcher and one of the teachers, I wanted to see the intervention succeed.

Ideally, multiple teachers would have used the rubrics to score each subjective portion and an average score would have been used.

How did the research impact the context?

Mia Jones, who taught three of the eight sections of the eighth grade history course, explained to a group of Independence Academy administrators that some of the differences she noticed between the control group and the experimental group were the kinds of things that might not show up after only six weeks on the types of assessments given during the study. Only one of her three eighth grade classes was a part of the experimental group, and she claimed that the discussions in that class were richer and more focused than those in her other two classes. She noticed that the students referred to specific texts more often and that they took pride in their ability to notice patterns and make connections between the different historical sources. This fits with what I observed when re-reading through the posttests for two classes, one from each group, after recording all of the scores and running the statistical analyses. In the one class that was a part of the control group, I noticed that only four out of the fourteen students in the class referred directly to the documents they analyzed in the HATs when answering the DBQ, while nine out of fifteen in the experimental group referred to the documents by name when answering their DBQs. The quality and length of the responses did not seem much different and both groups used about the same number of facts to support their claims, but the experimental group seemed to rely more heavily on the texts that were provided and more of them felt compelled to provide the sources of their evidence.

About a month after the school year ended, I presented the results and findings of the study to the interim head of school, middle school administrators, middle school learning specialist, and fellow eighth grade history teacher on June 24. These school leaders listened to

the forty-five-minute presentation and then asked questions and gave feedback. The biggest takeaway for several leaders who attended the presentation was that the students in the experimental group "still got the content." After the presentation, Assistant Head of Middle School Jason Newsome asked, "What's the harm?" If students in both groups learned the content, but one group had more opportunities to practice skills and analyze primary documents with expert guidance from the teacher, then he and the other administrators saw no reason to not try to use the DBL cycle more often. Interim Head of School Dr. Payton Stiles wondered if the intervention had a different effect on students with certain learning differences. He asked, based on an observation made by Eighth Grade History teacher Mia Jones, if this study or others like it revealed that the teaching methods used had a greater impact on students receiving academic accommodations and if those methods could be replicated in different grade levels or content areas.

Independence Academy Middle School is moving toward more vertical and horizontal alignment and is becoming more open to the idea of embracing written curricular standards, so I cannot be sure how much this study has contributed to some of the changes in the Middle School History Department. However, during the course of this study, the department adopted a set of national social studies standards and plans to evaluate how well the curriculum is already aligned with these standards in the 2019-2020 school year. After the evaluation period, the department plans to restructure the curriculum for each grade level to fill in content and skills gaps that the current teaching practices and plans do not address. During the meetings where the department discussed the different sets of standards, I was able to share some of what I learned from talking with teachers and curricular leaders in different divisions and from my research on historical thinking/reasoning skills.

Many students commented during and after the intervention that the assessments were too long and too difficult. Several of them did not believe the teachers when we told them that the pretest and parts of the posttest were just being used for evaluative purposes to see how much progress they made over the six-week period. Rumors began circulating through the grade that the assessment that we called a pretest was actually a high school history placement test. Students began asking the teachers if they could still enroll in the Honors World History I course if they did not perform well on the pretest. There is no way to determine how much student perceptions of the purpose of the tests affected student motivation, focus, and effort, but their comments suggest that the assessments may have caused some students to experience higher levels of stress and anxiety.

While the content pretest frightened, and discouraged some, others viewed it as a study tool. They commented, as we studied certain topics, "That was on the pretest!" Some even took notes in notebooks or on their laptop computers when they learned something that was covered on the pretest. We did not systematically track those behaviors so we cannot tell if those students performed better than their peers. However, the teachers did notice that there seemed to be some pedagogical benefit to giving the students a pretest. For the past four years, we have given students a list of key terms and content objectives at the beginning of each quarter and posted the terms for the week on our websites, but many students never looked at or utilized these lists. The pretest provided almost immediate feedback to the students about how much progress they needed to make to reach their academic goals in the course.

Summary of Results

Overall, there were very few statistically significant differences between the growth and achievement of the control and experimental groups before or after this intervention. On the

DBQ, students in the experimental group also outperformed the other students in terms of their ability to make plausible historical claims based on evidence from a variety of sources. While there was insufficient data to conclude that the intervention achieved its goals, there appears to be sufficient evidence to continue exploring the utility of the DBL Cycle and the RLH curriculum as additions to the Independence Academy Middle School history curriculum.

CHAPTER V

CONCLUSIONS

Summary of Chapter 4 Findings

The Independence Academy School's history courses are following the national trend in history education toward a greater emphasis on teaching, modeling, and assessing historical reasoning skills in addition to historical content knowledge. This trend has sparked greater interest in studying the ways that teachers, schools, districts, and higher education institutions measure success in history education. One of the goals of this study was to determine the effectiveness of a skills-based instructional intervention in eighth grade history classes at Independence Academy. The results of the study suggest that the two groups were nearly indistinguishable on every statistical measure of their performance and growth with very few exceptions.

The control group began the study with lower pretest scores on the content assessment, but finished with slightly higher scores. The ability of this group to close the tiny gap between them and the experimental group might suggest that over a longer period of time the difference between these two groups could have been even greater. One area where we observed a statistically significant difference was on the DBQ essay responses. Specifically, on the first category of the rubric--making claims, we noticed that experimental group members consistently made more specific claims that were supported by historical evidence from numerous sources. This was a little surprising because claim-making was not one of the target skills for the intervention. The intervention focused more specifically on Sourcing, Contextualization, Corroboration, and Close-Reading. However, we included claim-making because it was on the

ARCH Rubric, it is a big part of answering a DBQ, and it was a skill that we had taught and assessed on most major assessments throughout the school year even before beginning the study. While we did not intentionally focus on writing and making claims, we realized in realized that the RLH lessons required us to make claims about multiple documents in every class period and the guided questions included in these lessons ask students to make and defend claims, though they rarely, if ever, used the word "claim."

Discussion of Results in Relation to Extant Literature and Theories

According to Nokes (2011), corroboration and contextualization are difficult to teach and there are many barriers to observing any significant progress in these skills for adolescents over a short period of time. The results of this study fall in line somewhat with Nokes's observation and with the research of Reisman (2012) who studied the effects of RLH lessons in urban high school classes. Reisman observed that students in the treatment group outpaced their peers in close reading and sourcing skills, but neither group showed any real progress in their ability to contextualize or corroborate historical sources. Unlike the Reisman study, the eighth graders at Independence Academy performed similarly no matter what group they were in on assessments of close-reading and sourcing as well. There are many potential reasons for these differences between this study and Reisman's. One reason is that our assessment measured these skills in eighth graders instead of high school juniors. While many Independence Academy students read above grade level, their younger age must be considered when comparing their results to those of high schoolers.

Nokes (2011) contends that doing the intellectual work of historians' places cognitive demands on adolescents that they are not yet ready to handle. We saw that even the vocabulary in the questions caused some of the students' difficulty. While we did not conduct the Thinking

Aloud Protocols that similar studies have used, we did have several students express that their confusion about what the questions meant by terms like reliability and corroboration. During the lessons, we asked questions such as "Why do you trust or distrust this source?" (reliability) and "How does the information in this source support or contradict the information in the other source?" (corroboration). It might have been more useful and fairer to the students if we would have used the same language on the exam that we used in class.

One theme that showed up in the existing scholarship on Historical Reasoning was the notorious difficulty of measuring complex cognitive processes. DBQs in particular were criticized because they require students to go through so many cognitive and procedural steps in order to complete them that it is difficult to isolate and measure any particular skill because the researcher or evaluator does not know how other historical and non-historical skills and abilities interacted with one another to impact the results on the assessment of any particular skill (VanSledright, 2015). In order to answer the DBQs given as a part of this study, students at Independence Academy, were required to first analyze and answer questions about seven sources on the pretest and nine on the posttest. They had to understand the meaning of the question, formulate a historical claim, organize their arguments and evidence from multiple sources to support their claim, and then write a short essay. The researcher cannot be certain that the results on the different criteria on the DBQ rubric accurately represent the abilities of the test takers because other factors may have contributed more to their performance than placement in either one of the two groups. The HATs and WMCs were designed to better isolate individual skills than the DBQs while avoiding the problems with traditional multiple choice questions, however, our students struggled to make any noticeable progress on the WMCs, perhaps of difficulty with the wording of the documents, questions, or answers, and their progress on the HATs was

noticeable but not distinguishable between groups. Dr. Shaun Hutchins (June, 2019), in a personal online conversation, suggested that the assessment tools, with their one-, three-, and four-point scales, may not have provided a wide enough range of scores to identify noticeable differences between groups.

The research on historical thinking suggests that using multiple sources is superior to using a single source (J.D. Nokes, J.A. Dole, and D.J. Hacker, 2007). In this study, we used multiple sources with both groups. In order to determine the relationship between the instructional methods and the types of sources used, we would need to conduct the study with more than two groups. In our study the only independent variable we used was instructional method received. An even richer study would add number of sources used in class as an additional independent variable. Varying the number of sources used in the two groups might have created a greater difference, but the two teachers have never taught using a single source. In fact, for the course, the primary text, if it can be called that, is a collection of excerpts from a wide range of international textbooks called, *History Lessons: How Textbooks from Around the World Portray U.S. History.* Using a single traditional textbook or even a single source in each class would have gone against the nature of the course and would have been seen by the teachers, administrators, students, and others in the school community as an unnecessary step in the wrong direction.

Discussion of Lessons Learned

Since joining an independent school in 2011, I have not spent much time looking at and discussing the relationship between instruction and test performance. This study forced me to collect and analyze student performance data. One big lesson that I am taking away from this process is that I do not need to wait on our school to administer a standardized test in order to

track and respond to students' growth and achievement on the skills that I teach in my class. I do, however, need to develop an assessment and instructional plan for those skills or I will, like so many history teachers, overemphasize content knowledge acquisition at the detriment of teaching and assessing historical reasoning skills.

The HATs, WMCs, DBQs, and even simple recall questions about content all have a place in my classroom, but a big takeaway for me is that they do not all need to be used in the same lesson. The fact that I was unable to finish grading the pretest before the students took the posttest indicates that the assessment was too long to be helpful as practical instructional tool. Not only did some students lack the stamina or speed to finish the assessments, but I also did not have the time or the speed to get through the grading process in a timely manner. The discussions with students and other teachers about the different assessments were meaningful and helpful. Going forward, I plan to use these types of questions on formative and summative assessments, but I will give fewer test items at a time. Perhaps three to five HATs would be easier to grade than twelve or thirteen.

The fact that students showed so little improvement on the WMCs leads me to believe that the questions themselves were too difficult or the text excerpts were too complicated for them. Another problem with the WMCs in this study was that I used different ones for the pretest than for the posttest and I did not have a way to determine if the questions were of the same difficulty level. I do not know if it was even reasonable to expect the students to improve on the posttest because it is possible that the posttest was significantly harder. If it was, then any gains could have been canceled out by the increased difficulty. If I had used the same WMCs for the pretest and posttest, I would not have been able to tell if the improvements were the result of

the increased content knowledge, familiarity with the sources and the questions, or skills development.

I tried to use various test items to measure the effectiveness of an intervention at one school with one grade level. These quantitative tools were essential but insufficient for measuring effectiveness because much of what the lessons were designed to accomplish cannot necessarily be observed by looking at spreadsheets of students' scores on dozens of test items. After pouring over the performance data, I still want to know if students in the experimental group enjoyed the lessons more than their peers in the control group. I did not collect information on students' perceptions of the meaning and purpose of history. This would have been helpful for determining if the intervention changed the way students thought about history as a discipline. While their answers seem, on the surface to be similar in length and style, I would love to know if students in the experimental group made more connections between texts and if they are able to think, speak, and write more intelligently about primary and secondary documents.

The *Reading Like a Historian* lessons provided excellent examples of how to introduce students to the skills historians use without compromising the content that is so valuable to the study of the past. The Stanford History Education Group has not created a lesson about every historical topic that we cover in our eighth grade history course, but there are enough of them that we can use the lesson formats to create other lessons that give students the chance to learn and practice the same skills using different content. The RLH lessons identify the target skill(s) for each lesson. This was helpful because that determined the types of observations I made during think alouds with the students and the types of questions and activities that were included in the student handouts. This could become a new requirement within our department or at least

in my classroom, that each lesson have at least one target skill that is identified in the planning process, modeled during the lesson, practiced by the students, discussed by the class, and assessed during class or for homework.

I worked with several high school students during the summer right after finishing this intervention. I showed them several of the questions on the posttest. They breezed through the questions, even the ones that seemed most difficult to our eighth graders. About five of them looked at a HAT about the Baltimore Railroad Strike (UMBC, 2015) and immediately asked if the background information about the primary document was accurate and reliable (Sourcing). They then began discussing the dates listed on the document and what was happening in the United States at the time (Contextualization). Finally, they started reading line-by-line for clues about who was responsible for the damage during the riots (Close Reading). These students were just weeks removed from the A.P. United States History exam and three years of high school history courses, so their historical reasoning skills were fairly sharp. The ease with which they dispatched with tasks which seemed so daunting to my eighth grade students is an anecdotal sign that Independence Academy students do improve their historical reasoning and reading skills over time with practice. The conversations with these high schoolers also challenged me to give my eighth graders more of a head start on developing these skills by modeling, assessing, and tracking their progress on these skills beginning on the first day of eighth grade.

Implications for Practice

The use of Document Based Lessons during this study did not hinder students' ability to acquire historical content knowledge. This is perhaps the most important finding in the study because it suggests that history teachers at Independence Academy can feel free to dedicate class time to teaching, modeling, practicing, assessing, and tracking student development of historical

reasoning skills without worrying that they are sacrificing content and watering down their courses. While this study only observed eighth grade classes, there is no reason why DBLs and other lessons that explicitly and intentionally teach skills could not be used in other grade levels at Independence Academy.

The different types of assessments, especially the HATs, could provide teachers and students valuable tools for measuring progress on skills development. Teachers could use some of the HATs found on the Stanford History Education Group website and then develop their own to fit the target skills and content. Many of the HATs on the SHEG website use the same sentence and question stems, so teacher could easily swap out different documents and reword the questions. However, teachers of middle school students should definitely consider rewording the sentence and question stems to make the language more accessible to younger students. Tracking student progress on specific skills throughout the school year will allow teachers to make more informed instructional decisions and it will allow them to give students feedback about how they can improve their disciplinary skills and practices. Independence Academy teachers write report card comments for the first and third quarters of the school year. If history teachers use the instructional, assessment, and tracking practices used in this study, then they can include qualitative and quantitative data in those report comments and in other communication with students and parents.

As I mentioned in the Discussion of the Results, some students assumed that the pretest was actually an honors history placement exam. It was not. However, this was not a bad idea. Currently we recommend students for the World History I Honors course based upon their quiz, test, and semester exam averages. Moving forward, we could use a test similar to the ones given as a part of this study to help determine student readiness for the honors course. We could

continue to track the correlation between the placement test scores and student performance in the freshman history courses. We could also use this data in collaboration with the ninth grade teachers to guide our middle school history instruction because we can identify the skills that have the greatest impact on student success in high school history classes. If, for instance, we find that critical reading skills are the best predictor of success in high school, then we can target that skill more often in our middle school lessons.

For the past four years, our courses have typically covered one major topic each week. A takeaway from our study of Reconstruction, the Gilded Age, and the Great Depression was that these topics are too big to be covered in just one week. If the goal is to get students to think deeply about historical documents and content, then the students need more time to grapple with the sources and events of different time periods. Jumping to the next topic just for the sake of exposing students to more content and covering more chronological time is making the same mistake that studies like this one and good history instruction in general seek to correct. The advice I have for teachers like myself is: slow down! Or more accurately stated, "go deeper!" One reason why students made so little progress on their ability to contextualize sources is that they knew so little about what else was happening at the time. Students need time to explore, to investigate, to corroborate, and to organize their thoughts before they can develop well-informed opinions. This takes time.

Recommendations

As an internal study of a single subject, in a single grade level, at a single school, this study was not designed to provide generalizable results about the teaching of historical reasoning skills in other contexts. However, this study could serve as a starting point for a larger study in other independent and public schools. The assessments should be shorter so that they could be

easily completed within a single class period. An even better approach that testing students during their regular history class periods would be to test all students at a particular site at the same time so that the integrity of the questions and answers are not compromised by students sharing information about the assessments with students who have yet to take them. Rather than giving a single pretest and posttest, multiple assessments could be given to provide more data points and to determine if certain aspects of the intervention were more effective than others. Another major change that could be made to this study and future ones would be to lengthen the intervention period to see if the impact of the intervention increases over time.

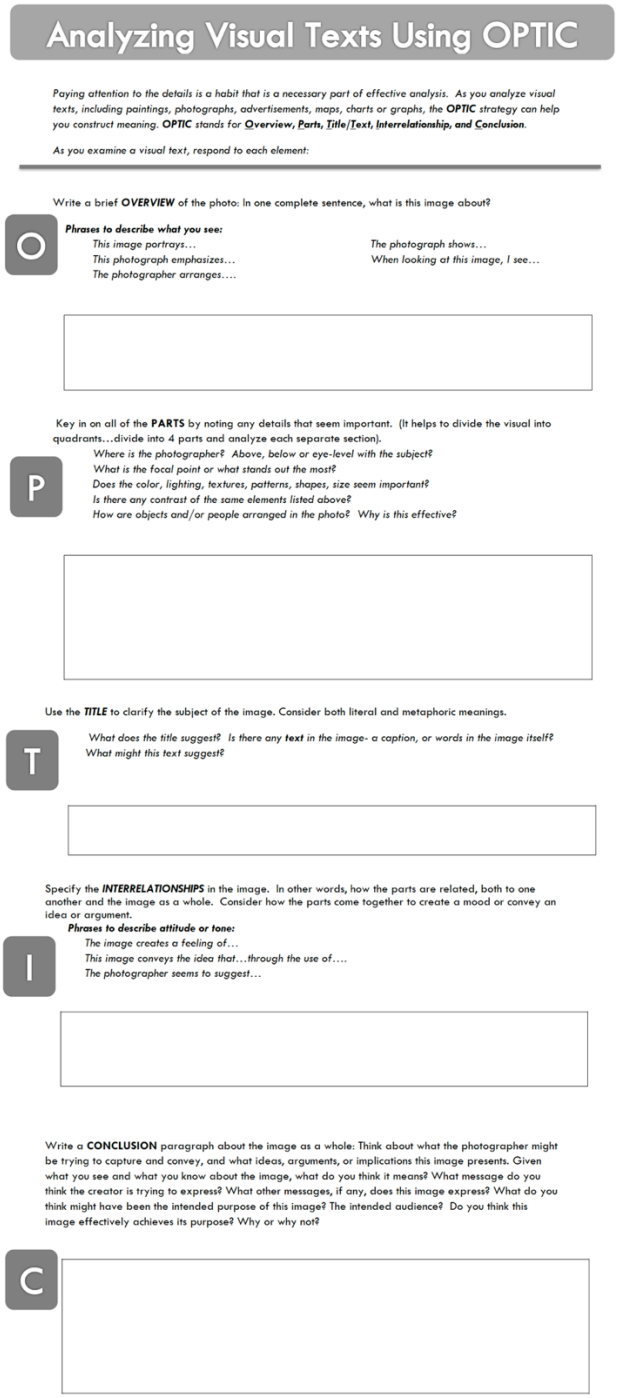
If I were to repeat this same study at Independence Academy, I would definitely include perception surveys, systematic classroom observations, analysis of student work portfolios, and student and teacher discussion groups. These more qualitative tools would help to compensate for the fact that not every change is measurable on a test item no matter how well it is designed.

One thing that this study did not investigate was the impact that the intervention had on different demographic populations. Are *Reading Like a Historian* lessons more effective with males than with females? A future study could analyze the impact of the intervention on English language learners, students with various learning differences, and students with different socioeconomic backgrounds. Public middle schools might be more interested in whether or not HATs and WMCs are any more informative and useful than the assessments they are already using. Researchers could use these assessments to see if performance on them is predictive of success on state-mandated standardized tests. Similarly, a school like Independence Academy, could track results on these assessments and see if there is a correlation to students' performance on AP history exams. If they find that they are not good predictors of success on exams, researchers could still look at the relationship between instructional and assessment strategies

and determine which formative and summative assessments will best help them monitor and improve students historical reasoning abilities.

The DBL model presented in this study is specific to the discipline of history, but a future study could observe the effectiveness of this model in other disciplines and determine if there are compound benefits to students having certain skills reinforced in different content areas. Students frequently analyze sources in science, English, and world language courses. Contextualization, sourcing, and critical reading could be just as important to those subjects as they are to the study of history. Of course, there are certain discipline specific skills and practices that different subject-area teachers might want to emphasize and model in their classrooms, but they could do this with certain types of sources using the DBL model.

Finally, at my school, we plan to incorporate the DBL model into more of our lessons in an attempt to see if prolonged exposure to these types of skills-specific lessons has a greater impact than the six-week intervention described in this document. Central to our approach will be the modeling of historical thinking and the interrogation of sources to discover their authorship, context, purpose, perspective, and reliability. We plan to provide students with different protocols for analyzing different types of sources. Figure 7 below gives an example of a protocol for analyzing visual sources that provides students with a series of sentence stems and questions that they would complete as they view an image. At the beginning of the year, the teachers would include the questions and sentence stems and would help them to create answers. As students gain confidence and skill, the teacher would remove the scaffolds and encourage students to generate their own questions and stems without teacher prompts.



Analyzing Visual Texts Using OPTIC

*Paying attention to the details is a habit that is a necessary part of effective analysis. As you analyze visual texts, including paintings, photographs, advertisements, maps, charts or graphs, the **OPTIC** strategy can help you construct meaning. **OPTIC** stands for **Overview, Parts, Title/Text, Interrelationship, and Conclusion**.*

As you examine a visual text, respond to each element:

Write a brief ***OVERVIEW*** of the photo: In one complete sentence, what is this image about?

O

Phrases to describe what you see:

This image portrays…
This photograph emphasizes…
The photographer arranges….
The photograph shows…
When looking at this image, I see…

Key in on all of the **PARTS** by noting any details that seem important. (It helps to divide the visual into quadrants…divide into 4 parts and analyze each separate section).

P

Where is the photographer? Above, below or eye-level with the subject?
What is the focal point or what stands out the most?
Does the color, lighting, textures, patterns, shapes, size seem important?
Is there any contrast of the same elements listed above?
How are objects and/or people arranged in the photo? Why is this effective?

Use the ***TITLE*** to clarify the subject of the image. Consider both literal and metaphoric meanings.

T

*What does the title suggest? Is there any **text** in the image- a caption, or words in the image itself? What might this text suggest?*

Specify the ***INTERRELATIONSHIPS*** in the image. In other words, how the parts are related, both to one another and the image as a whole. Consider how the parts come together to create a mood or convey an idea or argument.

I

Phrases to describe attitude or tone:

The image creates a feeling of…
This image conveys the idea that…through the use of….
The photographer seems to suggest…

Write a **CONCLUSION** paragraph about the image as a whole: Think about what the photographer might be trying to capture and convey, and what ideas, arguments, or implications this image presents. Given what you see and what you know about the image, what do you think it means? What message do you think the creator is trying to express? What other messages, if any, does this image express? What do you think might have been the intended purpose of this image? The intended audience? Do you think this image effectively achieves its purpose? Why or why not?

C

Figure 7 OPTIC Visual Analysis Protocol
(Harbor Teacher Preparation Academy, 2019). Reprinted.

Closing Thoughts

The study of history is more than just memorizing names and dates of famous people and major battles. History instruction and assessment need to catch up to this reality. This study attempted to add to the scholarship on historical reasoning by providing an example of an intervention and assessment approach that emphasized the acquisition of historical thinking skills. This study did not prove definitively that DBLs produce greater gains in terms of content knowledge and skills acquisition than other strategies for teaching history, but the results do suggest that over a short period of time the results are at least comparable. If the results are comparable, then it makes sense for history teachers to embrace the method that most resembles the work that professional historians actually do in their field.

While Sam Wineburg claimed that historical thinking was unnatural, I have come to believe that historical thinking is just as unnatural as algebraic or scientific thinking. Young children may naturally experiment with their environment, but they need to be taught the scientific process. Adolescents might think in terms of quantities and ratios, but they do not necessarily start developing formulas, proofs, and complex equations until prompted to do so by their math instructors. History is no different in that regard. Students study their past and are curious about what preceded them and how things came to be the way they are, but the importance of corroborating and sourcing documents may not be immediately obvious to the novice historian. Therefore, these skills must be taught, modeled, and reinforced in history classrooms. They must also be assessed. We need good tools to do that. Some studies have found HATs, WMCs, and DBQs to be more useful for assessing historical reasoning than traditional multiple choice questions. While WMCs are easiest to grade, the HAT appears to be one of the most practical assessment tools for measuring the development of historical reasoning

because they target specific skills, do not require much time to create, and can be graded relatively quickly when only a few are given at a time.

My coworker and fellow eighth grade teacher, Mia Jones argued that some of the benefits of teaching historical thinking may not show up on a standardized test or even a HAT or DBQ. Experts historians are not distinguished by their test scores. What sets them apart in the real world is their ability to write, to weave narratives, to interpret difficult texts, to corroborate sources, gather and make sense of evidence, and empathize with the historical figures as well as contemporary audiences. If the result of teaching historical reasoning is that students have deeper conversations about the meaning of historical texts and they learn to ask better questions of the sources they encounter, then the instructional investment in skills development was ultimately successful.

REFERENCES

Alexander, P. A. (2003). The development of expertise: The journey from acclimation to proficiency. *Educational Researcher, 32*(8), 10-14. doi:10.3102/0013189X032008010

American History Association, Stephens, H. M., Salmon, L. M., Haskins, C. H., Hart, A. B., Fox, G. L., Adams, H. B., McLaughlin, A. C. (1898). *The study of history in schools: A Report to the American Historical Association by the Committee of Seven.* The Macmillan Company; Macmillan and Co., Ltd. Retrieved from https://www.historians.org/about-aha-and-membership/aha-history-and-archives/historical-archives/the-study-of-history-in-schools

Appleby, J. (1995). Controversy over the national history standards. *OAH Magazine of History, 9*(3), 4. Retrieved from https://www.jstor.org/stable/25163026

Bell, J. C., & McCollum, D. F. (1917). A study of the attainments of pupils in United States history. *Journal of Educational Psychology, 8*(5), 257-274. doi:10.1037/h0074477

The Bradley Commission on History in Schools. (1989). Building a history curriculum: Guidelines for teaching history in schools. *The History Teacher, 23*(1), 7-35. doi:10.2307/494598

Buckingham, B. R. (1917). Correlation between ability to think and ability to remember, with special reference to United States history. Paper presented at the fifty-fifth annual meeting of the National Education Association of the United States, Portland, Oregon. Washington, D.C.: National Education Association's Secretary's Office.

Charap, L. G. (2015). Assessing historical thinking in the redesigned advanced placement United States history course and exam. In K. Ercikan, & P. Seixas (Eds.), *New directions in assessing historical thinking* (pp. 159-170). New York, NY: Routledge Ltd.

Claravall, E. B. (2017). Thinking like a historian: Developing disciplinary literacy in history among middle school struggling readers. *Literacy Today (2411-7862), 35*(1), 32-33. Retrieved from http://literacyworldwide.org

The College Board. (2017a). AP United States history: Course and exam description. Retrieved from https://secure-media.collegeboard.org/digitalServices/pdf/ap/ap-us-history-course-and-exam-description.pdf

The College Board. (2017b). Rubrics for AP histories: History disciplinary practices and reasoning skills. Retrieved from http://www.collegeboard.org

Commission on direction of the investigation of history and other social studies in the schools. (1930). *Journal of Educational Sociology, 3*(7), 437. doi: 10.2307/2961100

Cowgill, D., Armond, I., & Waring, S. M. (2017). Historical thinking: An evaluation of student and teacher ability to analyze sources. *Journal of Social Studies Education Research, 8*(1), 115-145. Retrieved from http://www.jsser.org

De La Paz, S. Ferretti, R., Wissinger, D., Yee, L., & MacArthur, C. (2012). Adolescents' disciplinary use of evidence, argumentative strategies, and organizational structure in writing about historical controversies. *Written Communication, 29*(4), 412-454. doi:10.1177/0741088312461591

Ercikan, K., & Seixas, P. (2015). Issues in designing assessments of historical thinking. *Theory into Practice, 54*(3), 255-262. doi:10.1080/00405841.2015.1044375

Ercikan, K., Seixas, P., Lyons-Thomas, J., & Gibson, L. (2015). Cognitive validity evidence for validating assessments of historical thinking. In K. Ercikan & P. Seixas (Eds.), *New directions in assessing historical thinking* (206-220). New York, NY: Routledge.

Frey, N., Fisher, D., & Hattie, J. (2017). Surface, deep, and transfer? considering the role of content literacy instructional strategies. *Journal of Adolescent and Adult Literacy, 60*(5), 567-575. doi:10.1002/jaal.576

Glaser, R., & Pittsburgh University, PA Learning Research and, Development Center. (1985). *Thoughts on expertise*. Retrieved from https://eric.ed.gov/contentdelivery/servlet/ERICServlet?accno=ED264301

Harper Teacher Preparation Academy (2019). "Analyzing visual texts using OPTIC." Retrieved from Harbor Teacher Preparation Academy https://harborteacherprep.com/ourpages/auto/2017/8/14/53164820/OPTIC.pdf

Huijgen, T., van Boxtel, C., van, d. G., & Holthuis, P. (2017). Toward historical perspective taking: Students' reasoning when contextualizing the actions of people in the past. *Theory and Research in Social Education, 45*(1), 110-144. doi: 10.1080/00933104.2016.1208597

Jacobs, H. H. (1997). Mapping the big picture: integrating curriculum and assessment, K-12. Alexandria, VA: Association for Supervision and Curriculum Development.

Kalisiki, P., Smith, K., & Huff, K. (2015). The importance of construct validity evidence in history assessment: What is often overlooked or misunderstood? In K. Ercikan, & P. Seixas (Eds.), *New directions in assessing historical thinking* (pp. 195-205). New York, NY: Routledge Ltd.

Kaliski, P., France, M., Huff, K., Thurber, A., & The College Board. (2011). *Using think-aloud interviews in evidence-centered assessment design for the AP world history exam* [PowerPoint slides]. Retrieved from https://eric.ed.gov/contentdelivery/servlet/ERICServlet?accno=ED563462

Kelley, T. L. (1914). *Educational guidance: an experimental study in the analysis and prediction of ability of high school pupils.* New York: Teachers College, Columbia University.

Kucan, L. (2017). Introducing the historical thinking practice of contextualizing to middle school students. *The Social Studies, 108*(5), 210-218. doi:10.1080/00377996.2017.1359483

Lazer, S. (2015). A large-scale assessment of historical reasoning and knowledge: NAEP U.S. history assessment. In K. Ercikan & P. Seixas (Eds.), *New directions in assessing historical thinking* (145-158). New York, NY: Routledge.

Lee, C. K. (2001). Anger, memory, and personal power: Fannie Lou Hamer and civil rights leadership. In B. Collier Thomas & V. P. Franklin (Eds.), pp. 139-170. *Sisters in the struggle: African-American women in the civil rights-black power movement.* New York, NY: New York University Press.

Lepore, J. (2018). *These truths: A history of the United States. first edition.* New York: W. W. Norton & Company, 2018

Lévesque, S. (2008). *Thinking historically: Educating students for the twenty-first century.* Toronto: University of Toronto Press.

Levisohn, J. (2017). Historical thinking--and its alleged unnaturalness. *Educational Philosophy and Theory, 49*(6), 618-630. doi:10.1080/00131857.2015.1101364

Levstik, L. S., & Barton, K. C. (2015). *Doing history: Investigating with children in elementary and middle schools*. New York, NY: Routledge. Retrieved from http://search.ebscohost.com.srv-proxy1.library.tamu.edu/login.aspx?direct=true&db=nlebk&AN=955904&site=ehost-live

Maina, N. S., & Montgomery County Public Schools (MCPS), Office of Shared Accountability. (2015). *Impact of advanced grade 8 U.S. history on participation and performance in advanced placement social studies courses in grade 9.* Montgomery County Public Schools. Retrieved from https://montgomeryschoolsmd.org

McCarty, R. (2017). *Leveraging historical thinking heuristics as warrants in historical argumentative writing.* (2017-05717-056). Retrieved from https://www.proquest.com/libraries/academic/dissertations-theses/

Méndez, S., & Tirado, F. (2016). Enhancing historical reasoning: A strategy including formative assessment with systematic continuous feedback. *International Journal of Educational Psychology, 5*(2), 187-219. Retrieved from https://hipatiapress.com/hpjournals/index.php/ijep/article/viewFile/1639/Mendez

Moore, J. R. 1., Alouf, J. L. 2., & Needham, J. (1984). Cognitive development and historical reasoning in social studies curriculum. *Theory & Research in Social Education, 12*(2), 49-64. doi:10.1080/00933104.1984.10505476

National history standards, part I standards in historical thinking. (1995). *OAH Magazine of History, 9*(3), 7-12.

New England History Teachers' Association Committee on, historical material, & Hazen, C. D. (1902). *Historical sources in schools.* The Macmillan company; Macmillan & co., ltd. Retrieved from http://ezproxy.library.tamu.edu/login?url=http://search.ebscohost.com.ezproxy.library.tamu.edu/login.aspx?direct=true&db=edshtl&AN=wu.89079737086&site=eds-live

Nokes, J. D. (2011). Recognizing and addressing the barriers to adolescents' "reading like historians". *The History Teacher,* (3), 379. Retrieved from https://thehistoryteacher.org/

Nokes, J. D., Dole, J. A., & Hacker, D. J. (2007). Teaching high school students to use heuristics while reading historical texts. *Journal of Educational Psychology, 99*(3), 492-504. doi: 10.1037/0022-0663.99.3.492

Peterson, M. M. (1994). *The road to Piney Point.* Piney Point Village, TX: Piney Point Village Historic Committee.

Reich, G. A. (2015). Measuring up? Multiple-choice questions. In K. Ercikan & P. Seixas (Eds.), *New directions in assessing historical thinking* (221-232). New York, NY: Routledge.

Reisman, A. (2015). The difficulty of assessing disciplinary historical reading. In K. Ercikan & P. Seixas (Eds.), *New directions in assessing historical thinking* (29-39). New York, NY: Routledge.

Reisman, A. (2012). Reading like a historian: A document-based history curriculum intervention in urban high schools. *Cognition and Instruction, 30*(1), 86-112. doi:10.1080/07370008.2011.634081

Reisman, A., & Wineburg, S. (2008). Teaching the skill of contextualizing in history. *Social Studies, 99*(5), 202-207. doi:10.3200/TSSS.99.5.202-207

Rouett, J., Britt, M. A., Mason, R. A., & Perfetti, C. A. (1996). Using multiple sources of evidence to reason about history. *Journal of Educational Psychology,* (3), 478. doi:10.1037/0022-0663.88.3.478

Samuelsson, J., & Wendell, J. (2016). Historical thinking about sources in the context of a standards-based curriculum: A Swedish case. *Curriculum Journal, 27*(4), 479-499. doi:10.1080/09585176.2016.1195275

Schuermann, P., & McGovern, M. (2016). Introduction to independent schools: Issues and opportunities. *Peabody Journal of Education, 91*(5), 569-573. doi:10.1080/0161956X.2016.1227160

Shanahan, T., & Shanahan, C. (2008). Teaching disciplinary literacy to adolescents: Rethinking content-area literacy. *Harvard Educational Review, 78*(1), 40-59. doi: 10.17763/haer.78.1.v62444321p602101

Smith, M. (2017). New multiple-choice measures of historical thinking: An investigation of cognitive validity. *Theory and Research in Social Education,* 1-34. doi:10.1080/00933104.2017.1351412

Smith, M., Breakstone, J., & Wineburg, S. (2018). History Assessments of Thinking: A Validity Study. *Cognition and Instruction. doi:* 10.1080/07370008.2018.1499646

Stanford History Education Group. (2018). Beyond the Bubble. Retrieved January 5, 2019, from https://sheg.stanford.edu/history-assessments

Stern, J. A. (2015). An appeal for sanity: A real-world look at the purpose, strengths, and weaknesses of the 2014 AP U.S. history framework. *Journal of the Gilded Age & Progressive Era, 14*(3), 400. doi: 10.1017/S1537781415000110

Stoel, G. L., van Drie, J. P., & van Boxtel, C. A. M. (2015). Teaching towards historical expertise. developing a pedagogy for fostering causal reasoning in history. *Journal of Curriculum Studies, 47*(1), 49-76. doi:10.1080/00220272.2014.968212

Strauss, V. (2018a). AP history course dropping centuries of history. *Washington Post.* Retrieved from https://www.washingtonpost.com/news/answer-sheet/wp/2018/06/15/ap-world-history-course-is-dropping-thousands-of-years-of-human-events-and-critics-are-furious/?utm_term=.8f0c0f67eafc

Strauss, V. (2018b). College Board restores 250 years to AP World History course after outcry over plan to cut 9,000 years. *Washington Post.* Retrieved from https://www.washingtonpost.com/news/answer-sheet/wp/2018/07/20/college-board-restores-250-years-to-ap-world-history-course-after-outcry-over-plan-to-cut-9000-years/?noredirect=on&utm_term=.70403271fb92

Straussman, S. (2015). A reflection on teaching a redesigned APUSH class. *Journal of the Gilded Age & Progressive Era, 14*(3), 419. doi: 10.1017/S1537781415000134

UMBC Center for History Education (2013). ARCH historical thinking skills rubric. Retrieved from https://www.umbc.edu/che/arch/images/ARCH_Historical_Thinking%20Skills_Rubric_Secondary_rev_2-17-14.pdf

Urdan, T. C. (2017). *Statistics in plain English. fourth edition.* New York: Routledge.

Van Drie, J. & Van Boxtel, C. (2008). Historical reasoning: Towards a framework for analyzing students' reasoning about the past. *Educational Psychology Review,* (2), 87. Doi: 10.1007/s10648-007-9056-1

VanSledright, B. (2015). Assessing for learning in the history classroom. In K. Ercikan, & P. Seixas (Eds.), *New directions in assessing historical thinking* (pp. 75-88). New York, NY: Routledge Ltd.

Wiggins, G. P., & McTighe, J. (2005). *Understanding by design*. Alexandria, VA: ASCD.

Wineburg, S. S. (1990). *Historical problem-solving: A study of the cognitive processes used in the evaluation of documentary and pictorial evidence*. doi: 10.1037/0022-0663.83.1.73

Wineburg, S. (1999). *Historical thinking and other unnatural acts*. *The Phi Delta Kappan,* (7), 488.

Wineburg, S. (2001). *Historical thinking and other unnatural acts: Charting the future of teaching the past.* Philadelphia: Temple University Press.

Wineburg, S. (2010). Thinking Like a Historian. *Teaching with Primary Sources Quarterly, 3*(1), 2-4.

Wineburg, S. (2007). Unnatural and essential: The nature of historical thinking. *Teaching History,* (129), 6-11. Retrieved from https://www.history.org.uk/publications/categories/teaching-history

Wineburg, S., Reisman, A., & Fogo, B. (2007). Historical evidence and evidence of learning. *International Journal of Social Education, 22*(1), 146-156. Retrieved from http://p2048-lib-ezproxy.tamu.edu.ezproxy.library.tamu.edu/login?url=http://search.ebscohost.com.ezproxy.library.tamu.edu/login.aspx?direct=true&db=eue&AN=507973078&site=eds-live

Wineburg, S., Smith, M., & Breakstone, J. (2018). What is Learned in College History Classes?, *Journal of American History*, *104*(4), 983–993. doi: 10.1093/jahist/jax434

APPENDIX A

QUARTER 1 HISTORY SKILLS ASSESSMENT PLAN

Standard	Skill	Measure 1	Measure 2	Measure 3	Measure 4
Chronological Thinking	1.Interpret data presented in timelines and create timelines	Pretest	Thematic **Online Timeline**	Colonization—**HAT**	MD*—**HAT**
	2.Reconstruct patterns of historical succession and duration to explain historical continuity and change.	Pretest	Colonization—**HAT**	Manifest Destiny —**DBQ**	**Online Paragraph** for Thematic Timeline
Historical Comprehension	3.Reconstruct the literal meaning of a historical passage	Pretest	Columbian Exchange —**WMC**	Jamestown Reading —**Guided Questions**	LA Purchase Reading —**Guided Questions**
	4.Identify the central question(s)	Pretest	Columbian Exchange —**WMC**	Jamestown Reading —**Guided Question**	LA Purchase Reading —**Guided Questions**
	5.Draw upon data in historical maps	Pretest	Reformation/Colonization —**Map Analysis**	MD —**Map Analysis**	Middle Passage — **Map Analysis**
	6. Utilize visual and mathematical data	Pretest	Columbian Exchange —**HAT**	Thanksgiving —**HAT**	American Progress Painting —**Image Analysis Protocol**
Historical Analysis and Interpretation	7. Analyze cause-and-effect relationships bearing in mind multiple causation including (a) the importance of the individual in history; (b) the influence of ideas,	Pretest	Columbian Exchange **—WMC**	Slavery in Southern Colonies —**WMC**	Proclamation Line Am. Rev. —**WMC**
	8. Draw comparisons across eras and regions in order to define enduring issues	Pretest	Jamestown v. Plymouth —**Exit Ticket**	N.E., Mid, S. Colonies —**WMC**	M.D. —**DBQ**
	9. Sourcing: Attribution: Cites all authors and all original dates of primary and secondary sources. Perspective: Evaluates the reliability sources based on the author's perspective and when and why they were produced.	Pretest	Comparing Columbus —**Paragraph**	Proc. 1763 + G. Washington Letter —**Guided Questions**	M.D. — **DBQ**

	10. Corroboration: Constructs an interpretation of events using information and perspectives in multiple sources. Identifies consistencies and inconsistencies among various accounts.	Pretest	Comparing Columbus —**Paragraph**	MD —**DBQ**	Movement Discussion —**Discussion Rubric**

*MD = Manifest Destiny

APPENDIX B

POINTS EARNED BY EACH STUDENT ON PRETEST CONTENT QUESTIONS

Gr.*	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
C	0	0	1	0	0	0	0	0	0.5	0	0	0.5	0	0	0	0	0	0	1	0
C	0	1	0	0	0	1	0	1	0	0	1	1	0.5	0	0	0	0	0.5	0	0
C	0	0	1	0	0	0	0	1	0.5	0	1	0	0	1	0	0	1	0	0	0
C	0	0	0.5	0	0	0	0	0.5	0.5	0	1	1	0	0	0	0	0	0	1	0
C	1	0	1	0	0	1	0	0	0.5	0	1	0.5	0	0	0	0	1	1	0	1
C	0	1	0	0	0	0	0	0.5	0	0	0	0.5	1	0	0	0	1	0.5	0	0
C	0	0	0	0	0	0	0	0	0.5	0	0.5	0.5	1	0	0	0	0	1	0	0
C	0	0	0	0	0	0	0	0.5	0	1	1	0.5	0	0	0	0	0	0.5	1	0
C	0	0	0	0	0	0	0	0	0.5	0	1	0	0	0	0	0	0	0	0	1
C	0	0.5	0	0	0	0	0	0.5	0	0	1	0.5	0	0	0		0.5	0	0	0
C	0.5	0	0	0	0	0	0.5	0	0	0	0	0.5	1	0	0	0	1	0	0	0
C	0.5	0	0	0	0	0	0	0.5	0	0	0	0.5	1	0	0	0	0	0	0	0
C	0	1	1	0	0	0	0	0.5	0	0	1	0.5	1	1	0	0	1	1	1	0.5
C	0	0	0	0	0	0.5	0	0.5	1	0	1	0.5	1	0	0	0	0	1	0	1
C	0	0	1	0	1	1	0	0	0	0	1	0	1	1	0	0	1	0.5	0	0.5
C	0	0	1	0	0	0	0	0.5	0	0	0	0	0	0	0	0	0	0.5	0	0
C	0	0	0.5	0	0	0	0	0.5	0	0	1	0.5	0	0	0	0	1	0	0	0
C	0	0	0	0	0	0	0	0	0	0	1	0.5	0.5	0	0	0	0	0.5	0	1
C	0	0	0.5	0	0	0.5	0	0	0	0	0	0	0	0	0	0	0.5	0	0	0
C	0	0	0	0	0	0	0	0.5	0	0	1	0.5	0	0	0	0	0	1	0	0
C	0	0	0.5	0	0	0	0	0.5	0	0	0.5	0	0	1	0	0	1	0	0	0
C	0	0	0	0	0	0.5	0	0	0	0	1	0.5	0.5	0	0	0	1	1	0.5	0
C	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0
C	0	0	1	0	0	0	0	1	0	0	0	0.5	1	0	0	0	1	1	0	1
C	0	0	1	0	0	0	0	0.5	0	0	0.5	0.5	0	0	0	0	0.5	0.5	0	0
C	0	0	2	0	0	0	0	1	0	0	1	0	0.5	0	0	0	0	0.5	1	0
C	0	0	1	0	0	0	0	0	0	0	1	0.5	0	0	0	0	0	0	0	0
C	0	0	0	0	0	0	0	0.5	0	0	0.5	0	0	1	0	0	0.5	1	1	0.5
C	0	0	0	0	0	0	0	1	0	0	1	0	1	1	0	0	1	1	1	1
C	0.5	0	0	0	0	0	0	0	0	0	0.5	0.5	0	0	0	0	0	0	0	0
C	0.5	0	0.5	0	0	1	0	0.5	0.5	0	1	0.5	1	0	0	0	1	0	0	0
C	0	0	0	0	0	0	0	0	0	0	0.5	0.5	0	1	0	0	1	0	0	0
C	0	0	0	0	0	0	0	0.5	0	0	0.5	0.5	0	0	0	0	0.5	1	0	0

C	0	0	0.5	0	0	0	0	0.5	0.5	0	1	0.5	0	1	0	0	1	1	0	1
C	0	0	0	0	0	1	0	1	0	0	1	0.5	0	1	0	0	0.5	1	1	0.5
C	0	0	0	0	0	0	0	0	0.5	0	1	0.5	0.5	0	0	0	0.5	1	0	1
C	0	0	1	0	0	0	0	0	0.5	0	1	0.5	1	1	0	0	1	0.5	0	0
C	0	0	0.5	0	0	0	0	0	0	0	1	0	0	1	0	0	1	1	0	1
C	0	0	1	0	0	0	0	0.5	0	0	0.5	0	0	0	0	0	1	0	1	0.5
C	1	0	0.5	0	0	1	0	0.5	0	0	1	0.5	0.5	0	0	0	1	1	0	0
E	0	0	1	0	0	0	0	0.5	0	0	1	0.5	0	0	0	0	1	0.5	0	0
E	0	0	0	0	0	0.5	0	0.5	0	0	0.5	0.5	0	0	0	0	0	0.5	0.5	0
E	0	0	0	0	0	0	0	1	0	0	1	0.5	1	0	0	0	1	0	0	0
E	0	0	0	0	0	0.5	0	0	0	0	0.5	0.5	0	0	0	0	1	0.5	0	0.5
E	0	0	0	0	0	0.5	0	0.5	0	0	0	0.5	0	0	0	0	0	0	0	0
E	0	0	0	0	0	0	0	0	0.5	0	1	0	0	0	0	0	0	0	0	0
E	0	0	0.5	0	0	0	0	0.5	0	0	0	0	0	0	0	0	0	0	0	0
E	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0.5	0	0	0
E	0	0	0.5	0	0	0	0	0.5	0	0	1	0	0	0	0	0	0	0.5	0	0.5
E	0.5	0	1	0	0	0.5	0	0.5	0	0	1	0.5	1	1	0	0	1	1	1	1
E	0	0	0.5	0	0	0.5	0	0.5	0.5	0	1	0.5	0	0	0	0	1	0	0.5	0
E	0	0	1	0	0	0.5	0	0.5	0	0	1	0	0	0	0	0	0.5	0	0.5	0.5
E	0	0	0	0	0	0	0	0	0.5	0	1	0.5	0	0	0	0	1	0	0	0
E	0	1	1	0	1	1	0	1	0.5	0	1	0.5	0	0	0	0	0	1	0.5	0
E	0.5	0	0	0	0	0	0	0	0	0	0	0.5	0	0	0	0	0	0	0	0
E	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
E	0	0	1	0	0	0	0	0.5	0.5	0	0.5	0.5	1	1	0	0	1	0.5	1	0.5
E	0	0	0	0	0	0	0	0	0	0	0.5	0.5	0	0	0	0	1	0	0	0
E	0	0	0.5	0	0	0	0	0.5	0	0	0	0	0	0	0	0	1	0	0	0
E	1	0	0.5	0	0	0	0.5	0.5	0.5	0	0	0.5	0	0	0	0	1	0	0	0
E	0	1	1	0	0	1	0	1	0	0	1	0.5	0	1	0	1	1	1	0	0
E	0	0	0	0	0	0	0	0.5	0	0	0	0.5	0	0	0	0	1	1	0.5	1
E	0	0	0	0	0	0	0	0	0	0	0	0.5	0	0	0	0	0.5	1	0	0
E	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1
E	0	1	1	0	0	0	0	1	1	0	1	0.5	0	0	0	0	0.5	1	0	0.5
E	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	1	0	0
E	0	0	0	0	0	0	0	0	0	0	1	0.5	0	1	0	0	1	1	0	1
E	0	1	0.5	0	0	0	0	0	1	0	1	0.5	1	1	0	0	1	1	1	0
E	0	0	0	0	0	0	0	0	0	0	1	1	0	1	0	0	1	0	0	0
E	0	0	0	0	0.5	1	0.5	0	1	0	1	0.5	0.5	1	0	0	1	0	1	1
E	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
E	0	0	1	0	0	0	0	0.5	0	0	1	0	0	0	0	0	0.5	1	1	0.5
E	0	1	0.5	0	0	0	0	1	0	0	1	0.5	0	0	0	0	0.5	0	0	0
E	0	0	0	0	0.5	0	0	1	1	0	1	0.5	1	0	0	0	1	1	1	0.5
E	0	0	0	0	0	0	0	0	0	0	1	0.5	0	0	0	0	1	0	0	0

E	0	0	1	0	0	1	0.5	1	0	0	1	0.5	1	1	0	0	1	1	0	0
E	0	0	1	0	0	0	0	1	1	0	1	0.5	1	0	0	0	1	1	0	1
E	0	1	1	0.5	0	0	0	0.5	0	0	1	0.5	0	0	0	0	1	1	1	0.5
E	0	0	0	0	0	0	0	0.5	0	0	0	0.5	0	0	0	0	0	0	0	0
E	0	0	0	0	0	0	0	0	0	0	0	0.5	0.5	1	0	0	0.5	1	0	0
E	0	0	0	0	0	0	0	0	0.5	0	0	0.5	0	0	0	0	0	0	0.5	0.5
E	0	1	0	0	1	0	0	1	0.5	0	1	0.5	1	0	0	0	1	1	1	0
E	0	1	1	0	0	0	0	1	0	0	1	0.5	1	1	0	0	1	1	0	1
E	0	0	0	0	0	0	0	0.5	0	0	0	0.5	0	1	0	0	1	0	0	1
E	1	1	1	0	0	1	0	1	0.5	0	1	0.5	1	1	0	0	1	0.5	1	1
E	0	0	1	0	0	0	0	0.5	0	0	1	0.5	0.5	0	0	0	0.5	0	0	0
E	1	1	0.5	1	0	0.5	0	0.5	0	0	1	0.5	0.5	1	0	0	1	1	1	0.5
E	0	0	0	0	0	0	0	1	0	0	1	0.5	0	0	0	0	0	0	0	0
E	0	0	0	0	0	0	0	0.5	0	0	0.5	0.5	1	1	0	0	1	1	0	1
E	0	0	0.5	0	0	0	0	0	0.5	0	0.5	0.5	1	1	0	0	1	1	0.5	0.5

*Gr. = Group, C = Control Group, E =Experimental Group

APPENDIX C

POINTS EARNED BY EACH STUDENT ON POSTTEST CONTENT QUESTIONS

Gr.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	sum
C	1	1	1	1	1	1	1	0	1	1	1	0	1	1	1	0	1	1	1	1	17
C	.5	1	1	1	1	1	.5	0	.5	0	1	0	1	1	.5	0	.5	1	1	1	13.5
C	1	1	1	1	1	.5	1	1	1	1	1	0	1	1	1	0	1	1	1	.5	17
C	1	0	1	1	1	.5	1	.5	1	0	1	0	1	1	1	0	.5	1	1	1	14.5
C	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	20
C	.5	1	1	1	.5	1	.5	.5	1	0	1	0	1	1	0	0	1	1	1	1	14
C	1	1	1	1	.5	1	1	.5	1	0	1	1	1	1	1	0	1	1	1	0	16
C	1	1	1	1	.5	1	.5	.5	1	1	1	0	0	1	1	0	.5	1	1	0	14
C	.5	1	0	1	1	1	.5	.5	.5	.5	1	1	0	1	1	0	1	1	1	1	14.5
C	1	1	1	1	1	.5	.5	.5	1	0	1	1	1	1	0	1	1	1	1	1	16.5
C	.5	1	.5	1	0	.5	1	.5	0	1	1	.5	1	1	0	0	1	1	1	.5	14
C	.5	1	1	.5	1	1	.5	1	1	0	1	1	1	1	1	0	1	1	.5	1	16
C	1	1	1	1	1	1	1	.5	1	1	1	0	1	1	1	0	1	1	1	1	17.5
C	1	1	1	1	1	1	1	1	1	.5	1	.5	1	1	1	0	.5	1	1	1	17.5
C	1	1	1	1	1	1	.5	1	1	1	1	0	1	1	1	1	1	1	1	0	17.5
C	0	1	1	1	1	1	1	.5	1	0	0	0	1	1	1	1	0	1	0	1	13.5
C	1	1	1	1	1	1	1	1	1	0	1	0	1	1	1	1	1	1	1	1	18
C	0	1	0	1	1	.5	1	.5	1	1	1	0	1	1	1	0	.5	1	1	.5	14
C	1	1	1	1	1	.5	.5	.5	1	0	1	.5	1	1	0	0	1	1	1	1	15
C	1	1	1	1	1	.5	1	.5	1	0	1	0	1	1	1	1	1	1	1	1	17
C	1	1	1	1	1	1	1	.5	1	.5	1	0	1	1	.5	1	1	1	1	1	17.5
C	1	1	.5	1	.5	.5	.5	1	1	1	1	0	.5	1	.5	1	1	1	0	1	15
C	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	19
C	1	1	1	1	1	1	1	.5	1	1	1	1	1	1	1	.5	1	1	1	1	19
C	.5	1	1	1	.5	1	1	1	1	0	1	0	1	1	1	1	1	1	1	1	17
C	1	1	1	1	.5	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	18.5
C	.5	1	1	1	1	1	.5	.5	1	1	0	1	1	1	0	0	1	1	1	1	15.5
C	.5	1	1	1	1	.5	1	.5	.5	.5	1	.5	1	1	0	0	1	0	1	1	14
C	1	1	1	1	0	1	1	0	1	1	1	1	1	1	.5	0	0	1	1	1	15.5
C	0	1	1	1	.5	1	0	0	0	.5	1	1	1	0	1	0	1	.5	1	.5	12
C	1	1	1	1	1	1	1	0	.5	1	1	.5	1	1	0	0	1	1	1	1	16
C	1	1	1	1	1	1	1	0	1	1	1	0	1	1	1	1	1	1	1	1	18.6
C	1	0	1	1	1	.5	1	.5	1	.5	1	1	1	1	0	0	1	1	1	.5	15

C	1	1	1	1	1	.5	.5	.5	1	0	1	0	1	1	0	0	1	1	1	1	14.5
C	1	1	1	.5	1	1	1	0	1	1	1	.5	1	1	1	0	1	1	1	.5	16.5
C	.5	1	1	1	1	0	.5	.5	1	0	1	1	1	1	0	0	1	1	1	1	14.5
C	1	1	1	1	1	1	1	0	1	1	1	.5	1	1	0	0	1	1	1	1	16.5
C	0	1	1	1	1	1	1	0	1	1	1	1	1	1	0	0	1	1	1	1	16
C	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	19
C	1	1	1	1	.5	1	1	.5	1	1	1	.5	1	1	1	1	1	1	1	1	18.5
E	1	1	1	1	1	.5	1	.5	.5	1	1	1	1	1	1	0	1	1	1	1	17.5
E	1	1	1	1	1	.5	.5	1	1	1	1	0	1	1	1	1	1	1	1	.5	17.5
E	1	1	.5	1	.5	1	.5	1	1	1	1	0	1	1	.5	1	1	1	1	1	17
E	.5	1	1	1	1	1	1	.5	1	.5	.5	0	1	1	1	1	1	1	1	1	17
E	1	1	1	0	1	1	0	1	1	1	1	0	1	1	.5	1	1	1	0	.5	15
E	0	1	1	.5	1	.5	.5	0	1	.5	1	0	1	1	1	1	1	1	1	1	15
E	.5	.5	1	1	.5	1	1	.5	.5	0	0	0	1	1	1	0	0	1	1	.5	12
E	1	1	1	1	1	.5	.5	1	1	0	1	.5	1	1	1	0	1	1	1	1	16.5
E	.5	1	1	1	1	.5	1	.5	1	.5	1	0	.5	1	.5	0	1	1	1	1	15
E	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	19
E	1	1	1	1	1	1	.5	1	1	1	1	0	0	1	1	.5	1	1	1	.5	16.5
E	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	20
E	.5	1	.5	1	1	1	.5	0	1	1	1	0	1	1	.5	0	1	1	1	1	15
E	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	20
E	0	1	1	0	1	1	.5	0	1	.5	1	.5	0	1	0	0	1	1	1	.5	12
E	1	1	.5	1	1	.5	1	0	1	1	1	0	1	1	1	0	1	1	1	.5	15.5
E	.5	1	1	1	1	1	.5	0	1	.5	1	1	1	1	0	0	1	1	1	1	15.5
E	.5	1	1	.5	0	1	.5	.5	1	0	0	0	0	0	.5	0	1	1	0	0	8.5
E	1	0	.5	1	1	1	.5	.5	.5	.5	1	0	1	1	0	0	1	1	1	.5	13
E	.5	1	1	.5	1	1	1	1	1	0	1	1	1	1	1	0	1	1	1	1	17
E	.5	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	19.5
E	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	0	1	1	1	1	18
E	0	1	0	0	0	.5	.5	0	1	.5	1	1	1	1	1	0	1	1	0	0	11
E	1	1	.5	1	0	.5	1	0	0	0	1	0	0	0	0	0	1	1	1	1	10
E	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	20
E	1	1	1	1	1	1	1	.5	1	1	1	1	1	1	1	0	1	1	1	1	18.5
E	1	1	1	1	1	1	1	0	1	.5	1	0	1	1	1	1	1	1	1	1	17.5
E	.5	1	1	1	1	1	1	0	1	1	1	0	1	1	.5	0	1	1	1	.5	15.5
E	1	1	1	1	1	0	.5	.5	1	1	1	1	1	1	0	0	1	1	1	1	16
E	1	1	1	1	1	.5	.5	0	1	0	1	0	1	1	0	0	1	1	1	1	14
E	1	1	1	1	1	.5	1	0	0	1	1	0	1	0	0	0	0	1	1	.5	12
E	1	1	1	.5	1	1	.5	.5	1	.5	1	.5	.5	1	0	.5	.5	.5	1	.5	14
E	1	0	1	1	1	.5	1	.5	1	1	1	0	1	1	1	.5	0	1	1	1	15.5
E	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	19
E	1	1	.5	0	.5	.5	.5	.5	1	.5	1	.5	1	1	0	0	1	1	1	1	14.5

E	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	.5	18.5
E	1	1	1	1	1	1	1	0	1	.5	1	0	1	1	1	0	1	1	1	1	16.5
E	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	0	1	1	1	.5	17.5
E	1	1	1	1	1	.5	1	0	1	.5	1	1	1	1	1	1	1	1	1	0	17
E	.5	1	.5	1	1	.5	1	1	1	.5	1	1	1	1	1	.5	1	1	1	1	17.5
E	1	1	1	1	1	.5	1	1	1	0	1	0	1	1	1	0	1	1	1	.5	16
E	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	0	1	1	1	1	18
E	1	1	1	1	1	1	.5	.5	1	1	1	1	1	1	1	1	1	1	1	1	19
E	.5	1	1	1	1	1	1	.5	1	0	1	.5	1	1	1	1	1	1	1	.5	17
E	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	19
E	1	1	1	.5	1	.5	.5	0	1	0	1	.5	1	1	1	0	1	1	1	1	15
E	1	1	1	1	1	1	1	1	1	1	1	.5	1	1	1	1	1	1	1	.5	19
E	1	0	0	1	1	0	.5	1	1	0	1	1	.5	1	1	0	.5	1	1	.5	13
E	.5	1	1	1	1	1	.5	0	1	1	1	.5	1	1	1	0	1	1	1	1	16.5
E	1	1	1	.5	1	1	1	.5	1	1	1	0	1	1	1	0	1	1	1	1	17

APPENDIX D

POINTS EARNED ON PRETEST WEIGHTED MULTIPLE CHOICE QUESTIONS

Group	1	2	3	4	5	6	7	8	9	10	11	12	13	WMC PRE AVG
Control	4	4	1	4	4	2	2	2	4	4	4	0	4	3
Control	4	4	2	4	2	1	2	0	4	4	2	4	4	2.846153846
Control	4	4	0	4	4	2	4	4	4	4	4	0	4	3.230769231
Control	2	4	0	1	4	4	4	4	4	4	2	0	4	2.846153846
Control	2	4	4	4	4	2	4	4	4	4	4	4	1	3.461538462
Control	4	4	1	4	4	2	4	4	4	2	2	2	1	2.923076923
Control	2	4	0	4	1	2	2	0	4	2	4	0	2	2.076923077
Control	2	1	2	2	1	1	2	0	4	2	2	0	2	1.615384615
Control	4	4	1	2	2	2	4	4	4	4	0	4	4	3
Control	4	4	0	1	4	2	4	4	4	4	0	0	4	2.692307692
Control	2	4	2	2	4	4	4	2	4	4	4	0	1	2.846153846
Control	4	4	1	4	4	4	4	0	0	4	4	0	0	2.538461538
Control	2	4	0	4	4	4	4	0	4	2	2	4	4	2.923076923
Control	2	4	0	4	1	2	4	0	4	4	4	4	2	2.692307692
Control	4	4	0	2	4	2	4	4	4	4	0	4	2	2.923076923
Control	4	4	1	4	4	2	4	4	4	2	2	2	1	2.923076923
Control	2	4	0	4	4	1	4	4	4	4	2	4	2	3
Control	2	4	0	4	1	2	2	0	4	2	4	0	2	2.076923077
Control	2	1	2	2	1	1	2	0	4	2	2	0	2	1.615384615
Control	4	4	1	2	2	2	4	4	4	4	0	4	4	3
Control	4	4	0	1	4	2	4	4	4	4	0	0	4	2.692307692
Control	4	4	1	4	4	4	4	0	0	4	4	0	0	2.538461538
Control	2	4	0	4	4	4	4	0	4	2	2	4	4	2.923076923
Control	2	4	0	4	1	2	4	0	4	4	4	4	2	2.692307692
Control	2	4	4	1	4	2	2	0	4	4	4	4	2	2.846153846
Control	4	4	0	2	4	2	4	4	4	4	0	4	2	2.923076923
Control	4	4	2	1	4	4	4	1	4	1	4	2	1	2.769230769
Control	4	4	0	1	4	2	4	4	4	4	4	4	1	3.076923077
Control	2	4	2	1	2	2	2	0	4	4	2	4	2	2.384615385
Control	2	4	4	4	2	1	0	4	0	2	0	4	0	2.454545455
Control	4	4	0	4	2	4	4	0	4	2	2	0	4	2.615384615
Control	4	4	4	4	4	2	4	1	4	4	0	0	1	2.769230769
Control	2	4	4	1	4	2	2	0	4	4	0	2	1	2.307692308

Control	2	4	1	1	4	4	4	4	4	4	2	2	0	2.769230769
Control	4	4	0	4	2	2	2	0	4	4	4	0	4	2.615384615
Control	4	4	1	4	2	2	4	4	4	4	4	2	0	3
Control	4	4	2	4	2	1	4	4	4	4	4	0	4	3.153846154
Control	4	4	4	2	4	2	4	4	4	4	2	0	4	3.230769231
Control	4	4	4	1	4	4	4	4	4	4	4	4	2	3.615384615
Control	2	4	4	4	4	2	4	4	4	0	4	0	2	2.923076923
Exp.	4	4	4	2	4	2	4	2	4	2	2	4	4	3.230769231
Exp.	4	4	4	4	1	4	4	4	4	4	2	4	1	3.384615385
Exp.	4	4	1	1	2	4	4	0	4	4	4	4	1	2.846153846
Exp.	4	4	2	2	0	LB	4	4	4	4	4	2	1	2.916666667
Exp.	4	4	0	4	2	2	2	0	2	4	4	4	1	2.538461538
Exp.	2	4	4	2	0	4	4	4	4	4	4	4	4	3.384615385
Exp.	1	4	4	1	4	2	2	0	4	4	0	0	LB	2.166666667
Exp.	4	4	4	2	4	2	4	2	4	4	2	4	2	3.230769231
Exp.	4	4	4	2	4	2	4	0	4	4	2	4	4	3.230769231
Exp.	4	4	2	2	4	1	2	4	4	4	1	0	2	2.615384615
Exp.	4	4	0	4	4	2	4	4	4	4	2	2	2	3.076923077
Exp.	4	4	2	4	1	4	4	4	4	4	4	0	1	3.076923077
Exp.	4	0	4	2	4	4	2	0	2	4	4	0	3	2.538461538
Exp.	2	4	4	2	4	2	4	4	4	4	4	4	4	3.538461538
Exp.	0	1	0	2	2	4	2	0	2	4	2	1	2	1.692307692
Exp.	2	4	0	1	0	2	4	2	4	4	4	4	1	2.461538462
Exp.	4	4	0	4	1	2	4	4	4	4	2	4	4	3.153846154
Exp.	0	0	4	0	2	4	1	4	2	4	4	2	4	2.583333333
Exp.	4	4	0	4	2	4	2	2	4	2	1	0	2	2.384615385
Exp.	4	4	1	2	4	4	4	4	4	2	0	4	1	2.923076923
Exp.	4	4	4	0	4	2	2	0	4	4	4	0	4	2.769230769
Exp.	4	4	1	4	4	2	4	4	4	4	0	4	1	3.076923077
Exp.	2	4	0	2	4	2	4	0	4	2	0	1	4	2.230769231
Exp.	4	4	4	C	4	4	4	4	4	4	2	1	2	3.416666667
Exp.	4	4	2	2	1	4	4	2	4	4	4	4	4	3.307692308
Exp.	4	4	0	4	4	2	2	0	4	4	4	2	2	2.769230769
Exp.	4	4	4	4	4	2	2	4	3	4	2	0	1	2.923076923
Exp.	4	4	1	4	1	2	4	2	4	4	1	0	1	2.461538462
Exp.	2	4	4	0	4	4	4	4	4	4	2	0	2	2.923076923
Exp.	4	4	0	4	2	4	4	4	4	4	4	4	2	3.384615385
Exp.	2	4	2	1	4	2	0	4	4	2	4	0	1	2.307692308
Exp.	4	4	0	4	4	2	4	4	4	4	4	0	4	3.230769231
Exp.	4	4	2	1	1	2	4	4	4	4	4	4	2	3.076923077
Exp.	4	4	2	4	4	2	4	4	4	4	4	4	1	3.461538462
Exp.	2	4	4	4	4	2	4	4	4	0	4	4	2	3.230769231

Exp.	4	4	2	2	0	0	4	4	4	4	4	4	4	3.636363636
Exp.	2	4	0	4	2	2	4	4	4	4	4	0	1	2.692307692
Exp.	2	4	4	4	2	1	4	4	4	4	4	4	0	3.153846154
Exp.	0	4	0	1	4	4	2	4	4	1	2	4	1	2.818181818
Exp.	2	4	0	4	4	4	4	4	4	2	2	4	1	3
Exp.	2	4	0	2	1	2	4	4	4	4	0	0	1	2.153846154
Exp.	4	4	0	2	4	2	2	2	4	4	4	4	1	2.846153846
Exp.	4	4	2	1	4	1	0	4	0	2	1	1	1	1.923076923
Exp.	4	4	1	1	4	4	4	4	2	0	1	0	0	2.230769231
Exp.	4	4	0	2	4	4	4	0	4	4	4	0	1	2.692307692
Exp.	2	4	0	4	4	2	4	4	4	4	2	0	1	2.692307692
Exp.	4	4	1	4	4	4	4	4	4	4	4	4	1	3.538461538
Exp.	4	4	0	1	2	4	0	4	4	4	2	0	2	2.384615385
Exp.	4	4	2	2	4	2	4	4	4	4	2	0	2	2.923076923
Exp.	4	4	4	4	2	4	2	0	4	4	4	0	4	3.076923077

APPENDIX E

POINTS EARNED ON POSTTEST WEIGHTED MULTIPLE CHOICE QUESTIONS

Group	1	2	3	4	5	6	7	8	9	10	11	12	13	WMC POST AVG
Control	4	4	4	4	4	1	4	4	4	0	4	4	4	3.461538462
Control	4	2	1	2	4	4	4	0	4	2	1	4	4	2.769230769
Control	2	1	0	4	2	4	4	4	4	4	2	4	1	2.769230769
Control	2	4	4	4	0	4	4	0	4	0	4	4	4	2.923076923
Control	4	4	0	4	4	4	4	4	4	4	4	4	4	3.692307692
Control	4	4	4	4	4	4	1	4	4	1	4	2	LB	3.333333333
Control	2	4	4	4	2	4	4	4	4	4	2	4	1	3.307692308
Control	4	2	2	4	4	4	1	4	4	4	4	4	1	3.230769231
Control	4	4	1	2	2	4	4	1	4	1	1	4	4	2.769230769
Control	4	0	4	0	4	4	0	1	4	1	4	4	1	2.384615385
Control	2	2	0	2	2	4	4	4	4	4	2	2	4	2.769230769
Control	4	4	4	4	4	4	1	4	4	4	4	2	2	3.461538462
Control	4	1	2	4	2	4	4	4	4	1	4	4	1	3
Control	1	4	1	2	2	4	1	4	4	4	4	2	1	2.615384615
Control	0	4	0	2	4	4	4	0	4	2	4	4	4	2.769230769
Control	2	4	4	0	4	4	0	4	4	1	4	4	2	2.846153846
Control	4	4	4	4	4	4	4	4	0	0	2	4	4	3.230769231
Control	2	4	2	0	4	1	1	4	4	1	4	4	1	2.461538462
Control	4	4	2	0	4	4	2	4	0	2	2	2	1	2.384615385
Control	4	2	0	2	4	2	1	4	0	2	0	4	4	2.230769231
Control	4	4	0	4	4	4	4	4	4	1	2	4	4	3.307692308
Control	2	2	2	0	2	2	4	0	4	2	2	4	2	2.153846154
Control	2	4	4	4	4	4	0	0	4	2	2	4	4	2.923076923
Control	2	2	4	4	1	4	4	4	1	1	0	4	4	2.692307692
Control	2	2	2	4	4	4	2	4	4	2	4	4	1	3
Control	4	4	1	4	4	4	2	4	4	4	4	4	4	3.615384615
Control	4	0	0	4	4	4	4	1	0	1	2	4	1	2.230769231
Control	2	4	4	4	4	1	4	4	4	2	2	4	2	3.153846154
Control	2	2	2	2	1	4	1	4	1	1	4	2	2	2.153846154
Control	2	4	4	2	4	4	1	0	4	1	4	2	1	2.538461538
Control	1	2	0	4	4	4	0	4	4	1	2	4	1	2.384615385
Control	4	4	4	4	4	4	1	4	0	2	4	2	4	3.153846154
Control	4	1	2	4	4	4	4	4	2	2	2	2	1	2.769230769

Control	2	2	2	1	0	4	4	4	4	4	4	2	4	2.846153846
Control	4	4	4	2	2	1	1	4	4	1	2	4	4	2.846153846
Control	4	4	2	4	4	4	1	4	0	1	0	2	2	2.461538462
Control	4	4	1	2	2	4	1	4	4	4	LB	LB	LB	3
Control	4	4	4	4	4	4	1	4	4	2	2	4	1	3.230769231
Control	4	4	4	4	4	4	1	4	4	4	4	2	1	3.384615385
Control	4	0	0	4	4	4	4	4	4	2	2	4	4	3.076923077
Exp.	4	4	4	4	2	4	4	4	1	2	4	2	2	3.153846154
Exp.	4	2	2	1	0	4	1	4	4	1	4	2	2	2.384615385
Exp.	4	2	0	4	4	4	4	0	4	0	4	4	2	2.769230769
Exp.	4	2	0	2	4	4	4	1	2	1	2	4	4	2.615384615
Exp.	4	4	4	2	2	1	2	4	4	2	2	4	2	2.846153846
Exp.	4	4	4	2	2	4	4	4	4	1	2	2	2	3
Exp.	1	2	2	4	0	0	2	4	4	2	2	4	2	2.230769231
Exp.	4	1	1	2	2	4	1	1	4	1	1	4	4	2.307692308
Exp.	4	4	4	2	4	1	1	4	0	0	4	4	4	2.769230769
Exp.	4	4	4	2	2	4	4	4	4	1	4	2	4	3.307692308
Exp.	4	4	1	4	2	4	2	0	0	1	4	4	4	2.615384615
Exp.	4	2	0	4	4	4	4	4	4	1	4	4	4	3.307692308
Exp.	1	2	4	4	4	4	4	4	4	2	4	4	2	3.307692308
Exp.	4	4	0	4	4	4	2	4	4	2	4	4	4	3.384615385
Exp.	2	0	2	0	4	4	1	4	4	4	4	2	2	2.538461538
Exp.	4	2	C	4	4	4	4	4	4	2	4	2	4	3.5
Exp.	2	2	2	4	4	4	1	0	4	4	4	4	4	3
Exp.	2	2	4	1	4	4	1	4	4	2	4	4	2	2.923076923
Exp.	2	1	0	4	4	4	4	4	1	2	4	4	1	2.692307692
Exp.	2	4	4	4	4	1	2	4	4	1	1	4	4	3
Exp.	4	4	4	0	1	4	4	4	4	4	4	4	2	3.307692308
Exp.	4	4	2	4	4	4	1	4	4	4	4	4	4	3.615384615
Exp.	2	4	0	4	2	4	1	4	4	4	4	4	2	3
Exp.	4	4	2	4	2	4	4	4	1	C	2	2	4	3.083333333
Exp.	4	2	0	2	2	4	2	4	4	2	1	4	2	2.538461538
Exp.	4	4	4	4	4	4	1	0	4	4	4	4	4	3.461538462
Exp.	2	4	2	2	4	4	4	4	4	1	2	4	1	2.923076923
Exp.	4	4	4	4	4	4	1	4	4	2	0	4	1	3.076923077
Exp.	2	0	2	4	4	4	1	4	4	4	4	4	1	2.923076923
Exp.	4	2	0	4	2	4	2	4	2	4	0	4	1	2.538461538
Exp.	4	4	4	2	2	4	0	0	0	1	0	2	2	1.923076923
Exp.	2	0	0	4	4	4	1	4	4	1	1	4	2	2.384615385
Exp.	2	4	1	4	4	4	4	4	0	4	4	2	2	3
Exp.	2	2	0	4	4	4	4	4	4	2	4	4	4	3.230769231
Exp.	1	2	0	4	1	4	4	4	4	2	4	2	2	2.615384615

Exp.	2	2	0	4	4	4	4	4	4	2	4	2	2	2.923076923
Exp.	2	2	1	2	2	4	1	1	4	1	2	2	1	1.923076923
Exp.	2	1	2	4	4	4	2	4	4	4	4	4	4	3.307692308
Exp.	4	4	1	4	4	1	0	4	4	1	1	2	2	2.727272727
Exp.	2	2	1	1	2	4	4	4	4	4	4	4	2	2.923076923
Exp.	4	4	1	4	2	4	4	4	4	2	4	4	1	3.230769231
Exp.	2	4	4	4	4	4	2	4	4	4	4	4	2	3.538461538
Exp.	4	1	0	4	2	4	4	4	4	2	1	4	1	2.692307692
Exp.	4	4	0	4	2	4	0	4	4	1	2	4	4	2.846153846
Exp.	2	4	1	4	0	4	1	4	4	4	4	4	4	3.076923077
Exp.	4	4	4	4	4	4	1	4	4	4	2	2	4	3.461538462
Exp.	2	4	4	4	2	4	4	4	4	4	4	4	4	3.692307692
Exp.	4	0	0	0	2	1	0	4	4	1	1	4	0	1.615384615
Exp.	2	4	1	4	4	4	1	4	4	4	4	4	2	3.230769231
Exp.	1	1	2	1	2	4	1	1	1	1	4	4	2	1.923076923

APPENDIX F

POINTS EARNED ON PRETEST HISTORY ASSESSMENTS OF THINKING

Group	1	2	3	4	5	6	7	8	9	10	11	12	HAT Avg
Control	3	3	3	3	3	3	1	3	1	3	2	3	2.583333333
Control	1	1	1	3	2	3	3	3	2	3	3	2	2.25
Control	1	1	2	3	3	2	1	2	1	3	3	3	2.083333333
Control	1	1	3	3	3	3	3	3	2	1	3	2	2.333333333
Control	2	1	3	3	3	3	1	3	2	3	3	3	2.5
Control	1	3	3	1	3	3	2	2	1	3	2	2	2.166666667
Control	3	1	3	1	3	3	3	3	1	3	3	1	2.333333333
Control	1	1	3	3	3	3	3	3	3	3	3	3	2.666666667
Control	2	1	3	1	2	1	1	1	1	1	1	2	1.416666667
Control	1	1	3	1	1	1	1	1	1	1	1	1	1.166666667
Control	3	1	1	1	3	3	1	3	1	3	1	1	1.833333333
Control	2	1	3	1	3	3	3	3	1	3	3	1	2.25
Control	1	3	3	3	3	3	3	3	2	1	3	3	2.583333333
Control	1	1	3	3	1	2	3	3	1	3	2	2	2.083333333
Control	1	1	3	3	3	3	1	1	3	1	3	1	2
Control	1	1	1	1	1	1	2	2	1	2	1	1	1.25
Control	1	1	3	1	1	1	1	1	1	3	3	1	1.5
Control	1	1	3	1	3	3	3	1	1	1	3	2	1.916666667
Control	3	1	3	3	1	1	3	1	1	2	3	1	1.916666667
Control	1	1	2	1	2	1	3	3	3	1	3	2	1.916666667
Control	1	3	3	1	3	3	3	3	1	3	1	2	2.25
Control	1	1	3	1	2	2	1	1	1	1	2	1	1.416666667
Control	2	1	3	3	2	2	3	3	3	1	3	3	2.416666667
Control	1	1	3	3	3	3	1	3	1	2	3	1	2.083333333
Control	3	1	1		3	3	3	3	3	2	3	1	2.363636364
Control	1	1	3	1	3	3	3	3	3	3	3	3	2.5
Control	2	1	3	1	3	3	2	3	3	1	3	1	2.166666667
Control	1	1	3	2	3	3	1	1	3	1	3	2	2
Control	1	1	2	2	3	3	1	3	2	1	2	2	1.916666667
Control	1	1	1	1	1	2	3	2	3	2	3	2	1.833333333
Control	2	1	3	3	1	2	1	1	3	3	3	1	2
Control	3	1	3	1	3	3	3	3	3	3	3	1	2.5
Control	3	3	1	3	3	3	3	3	3	3	2	3	2.75

Control	1	1	3	3	3	2	2	3	3	2	3	1	2.25
Control	1	2	3	3	3	3	3	3	3	2	3	1	2.5
Control	1	1	3	3	2	2	3	3	3	1	2	3	2.25
Control	1	1	3	1	3	3	1	3	1	1	3	1	1.833333333
Control	1	1	2	2	2	2	3	3	1	2	2	1	1.833333333
Control	2	3	3	3	3	3	3	2	3	3	3	3	2.833333333
Control	1	3	3	2	3	3	3	3	3	3	2	2	2.583333333
Exp.	1	1	3	1	2	2	3	3	1	2	3	3	2.083333333
Exp.	3	2	1	3	3	3	3	3	1	3	3	2	2.5
Exp.	1	1	2	3	3	3	3	3	1	3	3	3	2.416666667
Exp.	1	1	3	1	1	1	3	3	3	1	2	2	1.833333333
Exp.	1	3	3	3	2	3	3	3	2	1	1	1	2.166666667
Exp.	1	1	1	1	1	1	1	3	1	2	3	2	1.5
Exp.	1	1	3	1	3	3	1	2	2	1	2	1	1.75
Exp.	1	3	3	3	2	3	1	2	3	3	3	1	2.333333333
Exp.	1	1	3	1	3	3	3	3	1	3	2	2	2.166666667
Exp.	1	1	3	3	3	3	3	3	1	3	2	3	2.416666667
Exp.	1	1	3	3	3	3	3	3	1	1	3	2	2.25
Exp.	3	3	3	3	3	3	3	3	2	3	3	3	2.916666667
Exp.	1	2	1	3	3	3	1	3	1	3	2	2	2.083333333
Exp.	1	1	3	3	3	3	3	3	1	3	3	3	2.5
Exp.	3	1	3	1	2	1	1	1	1	1	3	1	1.583333333
Exp.	1	1	3	3	2	1	3	3	1	3	2	1	2
Exp.	1	1	3	3	3	3	1	2	1	1	1	3	1.916666667
Exp.	1	3	3	1	3	3	3	3	1	1	3	1	2.166666667
Exp.	3	1	3	1	3	3	2	3	1	2	3	1	2.166666667
Exp.	1	3	3	3	3	3	3	3	2	3	3	3	2.75
Exp.	1	1	3	3	2	3	3	3	1	3	3	1	2.25
Exp.	1	3	3	1	3	2	3	3	3	1	3	3	2.416666667
Exp.	1	2	1	1	1	1	3	3	1	1	3	1	1.583333333
Exp.	2	1	3	3	3	3	3	1	1	1	2	1	2
Exp.	3	1	3	3	3	3	3	3	2	3	3	3	2.75
Exp.	3	1	3	3	2	1	1	3	3	3	3	1	2.25
Exp.	1	1	3	3	3	3	3	3	2	3	3	1	2.416666667
Exp.	1	1	1	1	1	1	2	2	2	3	3	3	1.75
Exp.	3	1	3	3	3	3	3	3	2	3	3	2	2.666666667
Exp.	1	1	1	3	3	3	3	3	2	2	3	1	2.166666667
Exp.	1	1	2	1	1	1	1	1	2	2	3	1	1.416666667
Exp.	1	1	2	1	2	1	3	2	2	3	3	2	1.916666667
Exp.	1	1	3	1	3	3	2	2	2	1	3	3	2.083333333
Exp.	3	3	3	3	3	3	3	3	3	3	3	2	2.916666667
Exp.	1	1	1	1	3	3	2	2	2	1	1	1	1.583333333

Exp.	2	2	3	2	3	3	3	3	3	3	3	3	2.75
Exp.	1	3	3	2	3	3	3	3	2	1	3	3	2.5
Exp.	2	1	3	3	3	3	1	1	2	2	3	1	2.083333333
Exp.	1	1	3	1	3	3	3	1	2	1	3	1	1.916666667
Exp.	1	1	3	3	1	3	1	1	2	1	2	3	1.833333333
Exp.	3	1	3	3	1	1	1	2	1	3	3	1	1.916666667
Exp.	1	3	3	1	2	3	3	3	3	2	3	3	2.5
Exp.	3	1	3	3	3	3	3	3	1	3	3	3	2.666666667
Exp.	3	1	1	1	3	3	1	1	1	1	3	1	1.666666667
Exp.	3	1	3	3	3	3	3	3	1	3	1	2	2.416666667
Exp.	3	1	3	3	3	3	1	1	2	2	1	3	2.166666667
Exp.	3	3	3	3	3	3	3	3	3	3	3	3	3
Exp.	1	1	3	1	3	1	2	3	1	1	1	1	1.583333333
Exp.	3	1	3	1	1	1	1	3	1	3	2	2	1.833333333
Exp.	1	1	3	3	3	3	3	3	1	1	3	2	2.25

APPENDIX G

POINTS EARNED ON POSTTEST HISTORY ASSESSMENTS OF THINKING

Group	1	2	3	4	5	6	7	8	9	10	11	12	13	HAT Avg
Control	3	3	3	3	3	3	2	3	3	2	3	3	3	2.846153846
Control	1	1	3	1	3	2	1	3	3	2	1	2	2	1.923076923
Control	2	2	3	3	3	3	3	3	2	3	1	2	3	2.538461538
Control	3	3	3	3	3	3	2	3	2	2	3	3	3	2.769230769
Control	2	3	3	2	3	3	3	3	3	2	1	3	3	2.615384615
Control	3	2	3	3	3	3	3	1	1	2	1	3	2	2.307692308
Control	3	3	3	3	3	3	3	2	2	3	3	3	3	2.846153846
Control	3	3	3	3	3	3	3	1	1	2	3	3	3	2.615384615
Control	1	1	3	3	3	3	2	1	1	1	3	3	3	2.153846154
Control	2	3	3	3	3	3	2	1	3	2	1	2	3	2.384615385
Control	3	3	3	3	3	2	2	1	1	3	1	2	3	2.307692308
Control	3	3	3	3	3	3	2	3	3	3	1	3	3	2.769230769
Control	3	3	3	3	3	3	3	1	3	2	2	3	3	2.692307692
Control	3	3	3	3	3	3	3	3	2	2	3	3	1	2.692307692
Control	1	1	3	3	3	3	3							
Control	1	1	3	2	3	2	1	3	1	1	3	1	3	1.923076923
Control	3	2	3	1	3	3	3	3	2	3	1	3	3	2.538461538
Control	3	3	1	1	3	3	3	2	1	1	1	2	3	2.076923077
Control	3	3	1	2	2	3	3	2		3	1	3	3	2.416666667
Control	2	2	3	3	3	3	2	2	3	2	1	2	2	2.307692308
Control	3	2	3	3	3	3	3	2	3	2	3	3	3	2.769230769
Control	3	2	1	2	2	3	3	1	2	2	1	1	3	2
Control	3	3	3	3	3	3	3	3	3	3	1	3	3	2.846153846
Control	2	2	1	3	3	3	3	2	3	1	1	3	2	2.230769231
Control	3	2	3	3	3	3	3	3	1	2	2	3	3	2.615384615
Control	3	3	3	3	3	3	3	3	2	3	3	3	3	2.923076923
Control	3	2	3	1	3	3	3	2	1	3	1	3	3	2.384615385
Control	2	2	3	3	2	3	3	3	1	3	1	3	2	2.384615385
Control	2	2	3	3	3	3	3	2	1	1	3	3	3	2.461538462
Control	2	2	3	2	3	3	3	2	1	1	1	2	3	2.153846154
Control	3	3	3	3	3	3	3	2	1	1	1	3	2	2.384615385
Control	3	3	1	1	3	3	2	1	1	2	3	3	3	2.230769231
Control	1	1	3	3	3	3	3	2	1	1	1	3	3	2.153846154

Control	3	2	3	2	3	3	3	3	3	1	2	3	3	2.615384615
Control	2	2	3	3	3	3	3	3	3	1	1	2	2	2.384615385
Control	3	3	3	3	3	3	3	1	1	2	1	1	1	2.153846154
Control	3	2	3	3	3	3	2	2	1	2	1	1	1	2.076923077
Control	3	3	1	2	3	3	2	2	1	3	1	2	2	2.153846154
Control	3	3	3	3	3	3	3	2	3	3	3	3	3	2.923076923
Control	3	3	3	2	3	3	3	3		3	3	3	1	2.75
Exp.	2	2	3	2	3	3	3	3	2	3	3	3	3	2.692307692
Exp.	2	3	3	3	3	3	1	3	2	3	1	3	3	2.538461538
Exp.	2	1	3	2	3	3	1	1	2	2	1	2	3	2
Exp.	3	2	3	3	3	3	2	1	1	2	1	3	2	2.230769231
Exp.	2	2	3	3	3	3	3	2	1	1	3	3	3	2.461538462
Exp.	1	1	1	1	3	3	3	1	1	1	1	1	1	1.461538462
Exp.	3	3	3	3	3	3	3	1	2	1	3	1	3	2.461538462
Exp.	3	3	3	1	3	3	3	1	3	3	1	1	3	2.384615385
Exp.	2	3	3	3	3	3	3	1	1	1	1	2	3	2.230769231
Exp.	2	2	3	3	3	3	2	3	2	3	3	3	2	2.615384615
Exp.	3	3	3	3	3	3	2	1	1	2	3	3	2	2.461538462
Exp.	3	3	3	3	3	3	3	3	3	3	3	3	3	3
Exp.	3	3	3	3	3	3	3	2	1	2	1	3	1	2.384615385
Exp.	3	3	3	3	3	3	3	3	4	3	3	3	3	3.076923077
Exp.	1	3	3	3	3	3	2	2	1	1	1	3	2	2.153846154
Exp.	3	3	3	3	3	3	3	3	3	2	1	3	2	2.692307692
Exp.	3	3	3	3	3	3	3	3	1	2	1	3	3	2.615384615
Exp.	2	2	3	1	3	3	3	1	2	1	1	3	1	2
Exp.	3	3	3	3	3	3	2	1	2	1	3	3	3	2.538461538
Exp.	3	3	3	3	3	3	2	2	1	3	1	3	3	2.538461538
Exp.	2	3	3	1	3	3	3	3	3	3	1	3	3	2.615384615
Exp.	3	3	3	3	3	3	3	1	3	2	2	3	2	2.615384615
Exp.	3	3	3	3	3	3	2	1	1	1	3	3		2.416666667
Exp.	2	2	3	3	3	3	3			3	2	2	1	2.454545455
Exp.	3	3	3	3	3	3	3	3	2	2	1	2	2	2.538461538
Exp.	2	3	3	2	3	3	3	3	2	3	1	3	2	2.538461538
Exp.	2	3	3	2	3	3	3	3	2	3	1	3	2	2.538461538
Exp.	1	1	3	2	2	3	2	3	1	3	3	2	2	2.153846154
Exp.	3	3	3	3	3	3	3			3	2	3	3	2.909090909
Exp.	3	2	3	3	3	2	2	2	1	2	3	1	2	2.230769231
Exp.	3	3	3	3	3	3	3	3	1	3	3	3	3	2.846153846
Exp.	3	2	3	2	3	3	3	3	3	2	1	2		2.5
Exp.	2	2	3	3	3	3	3	1	1	2	3	3	2	2.384615385
Exp.	2	3	3	3	3	3	3	3	3	3	3	2	3	2.846153846
Exp.														

Exp.	4	4	4	4	4	4	4	2	3	3	2	3	3	3.384615385
Exp.	1	1	3	2	3	3	3	2	1	3	1	3	3	2.230769231
Exp.	1	1	3	3	3	3	3	1	3	3	3	3	2	2.461538462
Exp.	3	3	3	3	3	3	2	2	2	2	1	1	2	2.307692308
Exp.	2	3	3	2	1	3	1	1	1	3	3	2	2	2.076923077
Exp.	3	3	3	3	3	3	3	1	1	3	3	3	3	2.692307692
Exp.	2	3	3	3	2	3	3	2	3	3	1	3	3	2.615384615
Exp.	3	3	3	3	3	3	2	3	2	2	1	3	3	2.615384615
Exp.	3	3	3	3	1	3	2	1	1	1	1	1	1	1.846153846
Exp.	2	2	3	2	2	3	2	2	2	3	1	2	3	2.230769231
Exp.	3	3	3	1	3	3	3	2	1	1	3	1	3	2.307692308
Exp.	1	1	3	3	3	3	3	2	2	2	3	3	3	2.461538462
Exp.														
Exp.	3	3	3	3	3	3	3	3	3	3	2	3	3	2.923076923
Exp.	3	3	3	2	3	3	3	1	2	2	1	3	3	2.461538462

APPENDIX H

POINTS EARNED ON PRETEST DOCUMENT-BASED QUESTION

Group	Sourcing	Critical Reading	Corroboration	Context	Claim	Evidence	DBQ Avg
Control	2	2	2	3	3	3	2.5
Control	2	2	3	4	4	4	3.166666667
Control	2	2	2	2	4	3	2.5
Control	2	2	2	4	4	3	2.833333333
Control	3	2	3	4	4	4	3.333333333
Control	1	1	1	2	3	2	1.666666667
Control	1	1	1	1	3	3	1.666666667
Control	1	1	1	3	3	3	2
Control	1	2	2	2	3	2	2
Control	1	2	1	3	4	4	2.5
Control	1	1	1	2	4	2	1.833333333
Control	2	2	3	3	3	3	2.666666667
Control	3	2	4	4	4	4	3.5
Control	1	1	1	1	3	3	1.666666667
Control	1	1	1	2	3	2	1.666666667
Control	1	1	1	2	3	2	1.666666667
Control	2	2	2	3	4	2	2.5
Control	3	3	3	4	4	4	3.5
Control	1	1	2	2	3	2	1.833333333
Control	2	2	2	2	3	2	2.166666667
Control	2	2	2	3	3	2	2.333333333
Control	1	1	1	2	3	2	1.666666667
Control	3	3	3	4	4	4	3.5
Control	2	2	3	4	4	4	3.166666667
Control	3	3	4	4	4	4	3.666666667
Control	3	3	4	4	4	4	3.666666667
Control	1	1	1	2	2	2	1.5
Control	2	2	1	3	3	3	2.333333333
Control	1	1	1	2	3	3	1.833333333
Control	1	1	1	1	3	2	1.5
Control	1	1	1	3	4	3	2.166666667
Control	2	2	2	4	2	4	2.666666667
Control	1	1	1	3	4	3	2.166666667

Control	1	1	1	1	3	2	1.5
Control	1	2	3	4	4	4	3
Control	1	1	1	3	3	3	2
Control	1	1	1	2	2	2	1.5
Control	1	1	1	2	3	2	1.666666667
Control	2	2	1	4	3	4	2.666666667
Control	1	1	1	4	4	2	2.166666667
Exp.	2	1	1	2	4	2	2
Exp.	3	3	4	4	4	4	3.666666667
Exp.	1	2	2	2	4	4	2.5
Exp.	3	3	3	4	4	3	3.333333333
Exp.	2	2	2	4	4	3	2.833333333
Exp.	2	3	2	2	2	3	2.333333333
Exp.	2	1	1	3	3	3	2.166666667
Exp.	2	2	2	3	4	3	2.666666667
Exp.	1	2	1	2	4	3	2.166666667
Exp.	2	2	3	4	3	4	3
Exp.	2	1	2	2	4	2	2.166666667
Exp.	3	3	3	4	4	4	3.5
Exp.	1	1	1	3	3	3	2
Exp.	3	3	3	4	4	4	3.5
Exp.	2	2	1	2	2	3	2
Exp.	1	1	1	2	2	2	1.5
Exp.	2	3	3	4	4	4	3.333333333
Exp.	2	3	3	3	3	3	2.833333333
Exp.	2	3	1	2	2	2	2
Exp.	1	1	1	3	4	3	2.166666667
Exp.	1	2	1	3	4	3	2.333333333
Exp.	2	3	3	4	3	4	3.166666667
Exp.	1	1	1	1	4	2	1.666666667
Exp.	1	1	1	2	3	2	1.666666667
Exp.	1	1	1	1	4	2	1.666666667
Exp.	1	1	1	2	3	3	1.833333333
Exp.	2	2	2	2	1	2	1.833333333
Exp.	2	3	3	4	4	4	3.333333333
Exp.	1	1	1	4	3	3	2.166666667
Exp.	3	3	4	4	3	4	3.5
Exp.	1	1	1	2	1	2	1.333333333
Exp.	2	2	2	4	4	4	3
Exp.	1	1	1	1	4	2	1.666666667
Exp.	2	1	1	3	4	3	2.333333333
Exp.	1	1	1	2	2	2	1.5

Exp.	3	3	3	4	4	4	3.5
Exp.	2	2	2	1	2	2	1.833333333
Exp.	2	2	3	3	4	3	2.833333333
Exp.	1	1	1	2	3	2	1.666666667
Exp.	2	1	2	4	4	4	2.833333333
Exp.	2	2	2	3	3	4	2.666666667
Exp.	2	2	1	3	3	4	2.5
Exp.	1	2	1	4	3	3	2.333333333
Exp.	1	1	1	1	1	1	1
Exp.	2	2	1	4	4	4	2.833333333
Exp.	2	2	2	3	3	3	2.5
Exp.	1	2	1	2	4	4	2.333333333
Exp.	1	1	1	1	3	2	1.5
Exp.	2	2	1	2	2	2	1.833333333
Exp.	2	1	1	3	3	4	2.333333333

APPENDIX I

POINTS EARNED ON POSTTEST DOCUMENT-BASED QUESTION

Group	Sourcing	Critical Reading	Corroboration	Context	Claim	Evidence	DBQ Avg
Control	2	3	3	4	4	4	3.333333333
Control	2	3	2	4	4	4	3.166666667
Control	1	2	2	2	4	3	2.333333333
Control	1	1	1	2	3	3	1.833333333
Control	3	3	4	4	4	4	3.666666667
Control	1	1	1	3	2	3	1.833333333
Control	3	3	3	4	4	4	3.5
Control	2	2	2	2	3	3	2.333333333
Control	1	2	2	3	3	2	2.166666667
Control	2	2	2	4	3	4	2.833333333
Control	2	2	2	4	3	4	2.833333333
Control	3	2	3	4	4	4	3.333333333
Control	2	2	4	4	4	4	3.333333333
Control	1	2	2	2	4	4	2.5
Control	2	2	3	4	4	4	3.166666667
Control	3	3	3	4	4	4	3.5
Control	2	2	3	4	3	3	2.833333333
Control	3	3	3	4	4	3	3.333333333
Control	1	1	2	2	3	2	1.833333333
Control	1	2	1	2	3	2	1.833333333
Control	1	2	3	4	4	3	2.833333333
Control	1	1	1	2	3	2	1.666666667
Control	2	3	3	4	4	4	3.333333333
Control	1	1	1	3	3	2	1.833333333
Control	2	2	1	2	3	2	2
Control	1	3	4	4	4	4	3.333333333
Control	3	2	2	4	4	4	3.166666667
Control	1	1	1	3	4	4	2.333333333
Control	1	2	2	4	2	4	2.5
Control	1	1	1	3	2	2	1.666666667
Control	3	2	4	4	4	4	3.5
Control	3	3	4	4	4	4	3.666666667
Control	2	2	3	4	4	4	3.166666667

Control	1	1	1	2	3	2	1.666666667
Control	1	1	1	3	3	2	1.833333333
Control	2	2	3	3	3	3	2.666666667
Control	1	1	1	2	2	1	1.333333333
Control	1	1	1	3	4	3	2.166666667
Control	2	2	3	4	4	4	3.166666667
Control	1	1	2	2	3	2	1.833333333
Exp.	3	2	1	4	3	4	2.833333333
Exp.	2	2	4	4	4	4	3.333333333
Exp.	2	1	2	3	4	3	2.5
Exp.	2	3	3	3	4	4	3.166666667
Exp.	2	3	3	4	4	4	3.333333333
Exp.	1	1	1	2	3	2	1.666666667
Exp.	1	1	1	2	4	4	2.166666667
Exp.	2	3	3	4	4	4	3.333333333
Exp.	2	1	2	3	4	2	2.333333333
Exp.	2	2	2	3	3	3	2.5
Exp.	3	3	4	4	4	4	3.666666667
Exp.	1	1	1	4	3	4	2.333333333
Exp.	2	2	4	4	4	4	3.333333333
Exp.	1	3	3	4	4	2	2.833333333
Exp.	2	1	3	4	4	3	2.833333333
Exp.	2	2	3	4	3	3	2.833333333
Exp.	4	4	4	4	3	4	3.833333333
Exp.	1	1	1	3	4	2	2
Exp.	2	2	3	4	4	4	3.166666667
Exp.	3	3	4	4	4	4	3.666666667
Exp.	2	2	2	4	4	4	3
Exp.	2	2	1	3	4	3	2.5
Exp.	1	2	3	3	3	3	2.5
Exp.	1	1	1	2	2	2	1.5
Exp.	3	3	3	4	4	4	3.5
Exp.	3	2	3	4	4	4	3.333333333
Exp.	2	2	4	4	4	4	3.333333333
Exp.	1	1	1	2	4	2	1.833333333
Exp.	3	2	3	4	4	4	3.333333333
Exp.	1	1	1	2	4	2	1.833333333
Exp.	3	2	4	3	4	4	3.333333333
Exp.	3	3	4	4	4	4	3.666666667
Exp.	2	2	2	4	4	3	3
Exp.	3	2	4	4	4	4	3.5
Exp.	2	3	3	4	4	4	3.333333333
Exp.	3	4	2	4	4	4	3.5

Exp.	1	1	1	3	4	3	2.166666667
Exp.	1	2	1	3	4	3	2.333333333
Exp.	2	2	2	3	4	4	2.833333333
Exp.	1	2	2	3	4	3	2.5
Exp.	1	1	1	3	2	2	1.666666667
Exp.	2	2	3	4	4	4	3.166666667
Exp.	1	1	1	2	4	2	1.833333333
Exp.	3	2	3	4	4	4	3.333333333
Exp.	2	2	2	4	3	4	2.833333333
Exp.	2	2	2	4	4	4	3