

The Three-Legged Race: Exploring the Relationship between History and Social Studies Teaching and Standardized Tests

Author: Dianna Lynn Gahlsdorf Terrell

Persistent link: <http://hdl.handle.net/2345/1104>

This work is posted on [eScholarship@BC](#),
Boston College University Libraries.

Boston College Electronic Thesis or Dissertation, 2010

Copyright is held by the author, with all rights reserved, unless otherwise noted.

BOSTON COLLEGE
Lynch School of Education

Department of
Teacher Education, Special Education, Curriculum and Instruction

THE THREE-LEGGED RACE: EXPLORING THE RELATIONSHIP BETWEEN
HISTORY AND SOCIAL STUDIES TEACHING AND STANDARDIZED TESTS

Dissertation
By

DIANNA GAHLSDORF TERRELL

Submitted in partial fulfillment
Of the requirements for the degree of
Doctor of Philosophy

May 2010

© Copyright by DIANNA GAHLSDORF TERRELL
2009

ABSTRACT

THE THREE-LEGGED RACE: EXPLORING THE RELATIONSHIP BETWEEN HISTORY AND SOCIAL STUDIES TEACHING AND STANDARDIZED TESTS

By Dianna Gahlsdorf Terrell

Marilyn Cochran-Smith, Ph.D., Chair

A healthy democratic society requires citizens with both the knowledge to understand the problems it faces and the dispositions to solve them. Yet recent studies have shown that citizens in the United States are losing the democratic habits required to solve social problems. Moreover, results on standardized assessments in United States history including the National Assessment of Educational Progress bear out the fact that the historical knowledge of typical American high school graduates is woefully lacking (Gaudelli, 2002; Shenkman, 2008). Some blame teachers for failing to teach students meaningful content, and others counter that students' poor performance signals a problem with the test's construction rather than with teachers. This dissertation was designed to inform the debate through a systematic study of the orientations of history and social studies teachers in Massachusetts, the skills and constructs measured by the MCAS-US history test, and the relationship between the two.

This study considered the complex relationship between teachers' orientations and the skills and constructs measured on the MCAS-US test via two research designs. First, a survey of Massachusetts history and social studies teachers was conducted to analyze the orientations from which teachers approach the subject. Second, a content analysis of the MCAS-US test was conducted to identify the skills and constructs

assessed on the test. Both the survey and the content analysis were carried out through the theoretical lens of democratic pragmatism, and both employed the same framework for understanding the varied ways that history and social studies is taught. Findings point to a very clear misalignment between orientations of history and social studies teachers and the skills and constructs measured by the MCAS-US test. This conjures up an image of a three-legged race where the two participants appear to work against one another. The dissertation concludes with a discussion of the implications of the study, including ways that test developers and history and social studies teachers can make progress toward the shared goal of improving civic knowledge and participation.

ACKNOWLEDGEMENTS

In the many years it has taken me to complete this doctoral program, it seems that an increasing number of people have asked how I sustain the energy to write a dissertation, work as an assistant professor, raise a young daughter and prepare to welcome a second daughter. Without hesitation, I tell these people that, as is true of raising a child, writing a dissertation takes something of a village. It is likely that I would have given up long ago on this path had it not been for the incredibly supportive network of extraordinary people I have been lucky to work with over the last several years.

First, to Dr. Marilyn Cochran-Smith, who as an exceptional dissertation chair, personal, academic and professional advisor never failed to uphold anything but the highest expectations for my thinking and writing, there really are no words to express my gratitude for your ability to direct, reflect on and critique this work with deep and thorough attention. I remain in awe of your ability to move between and among your many roles as an academic, an intellectual and a fine individual with steadfast focus and dedication. (Your four o'clock in the morning email responses leap to mind.) You are a rare example of an individual who can talk the talk and walk the walk. Thank you.

To Dr. Joe Pedulla, who as my methodological advisor and committee member was able to keep me on the straight and narrow path of proper statistics, thank you. It was not long before I began to see why so many have spoken admiringly of your teaching ability in such a panic-inducing subject. For all of the anxious situations I confronted with this work, you were able to offer correction and guidance, and, like a quality teacher often does, you always stopped far short of providing the answers. Your subtle approach reminded me throughout that the doctoral dissertation should be a learning experience.

To Dr. Pat McQuillan, who as a fellow history and social studies advocate kept me questioning the “best methods” of teaching the subject we have both made our lives’ work, thank you. Your unique perspectives of classroom teaching inspired one of the largest critiques of traditional theory in history and social studies education that this study offers. Our impromptu conversations in the halls of Campion often reminded me of the timeliness and practical applicability of my study to a real and pressing policy issue.

To the many doctoral students at Boston College with whom I have had the privilege of working, both in class and on projects associated with the Teachers for a New Era Qualitative Case Studies group, thank you for setting a high intellectual bar and for sustaining the collegial atmosphere that sets Boston College apart from other graduate programs. I have developed lasting friendships with many of you which I, no doubt, will cherish and draw strength from as we transition from classmates to colleagues in the years to come.

To the Terrell family and Gahlsdorf family (including the ones who have since changed those names for others), you may have often revealed that you never quite knew exactly

what I was doing. Nevertheless, you continuously expressed your pride and support. Bearing witness to your life events – and having you bear witness to my own – has been my constant reminder that life moves on at a remarkable pace whether or not my head is buried in books. It has been these personal events, not necessarily professional achievements, which have kept me grounded, sane and (dare I say) personable over the last several years.

To my daughter, Zoë, who never fails to make me laugh out loud several times a day, and to the daughter I have yet to meet but who has been with me these last eight months, thank you both for providing an incentive to finish this work and for giving me something to look forward to once it is done. I now have something to point to as evidence that intellectual curiosity is personally rewarding, and that the loftiest ambitions you hold can be realized if only you commit to taking the first step. I look forward with great anticipation to helping you define and achieve your own ambitions in life.

And to my wonderful husband, Rien, who taught me the importance of committing to that first step, I would not be where I am today were it not for your simple suggestion, “Just apply, Dianna.” Your balance, humor, serenity, intellect and steadfast support have been an inspiration. You have been my constant companion, my first reader on nearly every draft, and the best friend I could ever hope for. Thank you for understanding how important this was for me and for being just as committed to it as I have been. So, what’s next?

TABLE OF CONTENTS

ABSTRACT.....	I
ACKNOWLEDGEMENTS.....	III
TABLE OF CONTENTS	V
LIST OF TABLES, LIST OF FIGURES, LIST OF SAMPLE QUESTIONS	VIII
CHAPTER ONE: BACKGROUND AND PROBLEM	1
The Problem of Diminishing Democratic Knowledge and Participation	2
Addressing the Problem: A Call for Improved Civic Education	11
Massachusetts: A State on the Brink of Civic Education Reform	15
Organization of the Dissertation	16
CHAPTER TWO: FRAMING THE STUDY	22
The Need for a Theory: Democratic Pragmatism	23
<i>Democratic Pragmatism and the Role of Public Schools.....</i>	<i>27</i>
<i>Democratic Pragmatism on What Democratic Education Should Look Like.....</i>	<i>29</i>
<i>Concluding Thoughts: How Democratic Pragmatism will inform this study</i>	<i>32</i>
Related Literature on Curriculum Orientations in HSS.....	35
<i>Conceptual Studies: Gatekeeping and Orientations in the Field.....</i>	<i>37</i>
<i>The “Three Traditions” Thesis as a Framework for Understanding Orientations ..</i>	<i>40</i>
<i>Findings of Quantitative Empirical Studies: Demographic Differences.....</i>	<i>57</i>
<i>Findings from Qualitative Empirical Studies: Teachers as Gatekeepers</i>	<i>62</i>
<i>Concluding Thoughts on History and Social Studies Teachers’ Orientations in the Field.....</i>	<i>65</i>
Related Literature on Standards and Assessments in HSS	68
<i>Conceptual literature on standards and assessments in HSS.....</i>	<i>69</i>
<i>Empirical studies on standards and assessments</i>	<i>90</i>
<i>Concluding Thoughts on Standards and Assessments in History and Social Studies</i>	<i>111</i>
CHAPTER THREE: RESEARCH DESIGN	114
Design I: Survey of History/Social Studies Teachers.....	116

<i>Data Collection</i>	116
<i>Data Analysis</i>	130
Design II: Massachusetts History and Social Studies Standards and Assessments....	134
<i>Data Collection</i>	134
<i>Data Analysis (Content Analysis)</i>	136
CHAPTER FOUR: THE NUANCES OF MASSACHUSETTS HSS TEACHERS	142
Analysis of the Survey Instrument: Five or Three Orientations of HSS Teachers? ...	143
<i>Scale Reliability Analysis</i>	144
<i>Repeated Measures ANOVA Test</i>	152
<i>Factor Analysis</i>	155
<i>Measuring and Understanding Teachers' HSS Orientations</i>	167
The Prevalence of the Revolutionary Spirited RI Orientation in Massachusetts.....	174
<i>Reformers and Reflective Inquirers: Massachusetts HSS Teachers</i>	174
<i>True Blue Teachers in a Blue State: Demographically Distinct from US teachers</i>	183
Massachusetts Teachers Seem to Work from a Theory.....	192
CHAPTER FIVE: A STATE US HISTORY TEST WITHOUT A PATH FORWARD.....	204
Process for Content Analysis.....	206
Findings from Content Analysis.....	218
Discussion of the Content Analysis Findings.....	232
<i>Multiple Choice Items in the MCAS-US Test</i>	233
<i>The American Grand Narrative According to MCAS-US Test</i>	238
Analysis of the Relationship between HSS Teachers' Orientations and the Concepts and Skills Measured by the MCAS-US Test	244
<i>Is the Content on the MCAS-US Test Consistent with Teachers Orientations?</i>	246
<i>How Teachers Viewed State Standards and the MCAS-US Test</i>	254
CHAPTER SIX: IMPLICATIONS OF THIS STUDY	268
Out of Alignment: The MCAS-US Test and Massachusetts HSS Teachers	269
<i>In Search of a Guiding Theory for the MCAS-US Test</i>	276
<i>Massachusetts Teachers are Working from a Theory</i>	288
<i>Out of Alignment: Should the Test or Teachers come in Line?</i>	291
BIBLIOGRAPHY	299

APPENDIX A: WORK PLAN.....	319
APPENDIX B: ORIGINAL VINSON SURVEY INSTRUMENT	321
APPENDIX C: REVISED SURVEY INSTRUMENT USED IN THIS STUDY.....	327
APPENDIX D: INVITATION TO PARTICIPATE IN VOLUNTARY RESEARCH	332
APPENDIX E: CONTENT ANALYSIS MATRIX.....	333
APPENDIX F: SAMPLE QUESTIONS FROM THE MCAS-US TEST	334
NOTES	350

LIST OF TABLES, LIST OF FIGURES, LIST OF SAMPLE QUESTIONS

Tables

Table 1: <i>Breakdown of the Instrument by Item Type</i>	123
Table 2: <i>Proposed Research Questions and Analysis Techniques</i>	131
Table 3: <i>Scale Reliability Analysis of Items Measuring Five Orientations</i>	145
Table 4: <i>Scale Reliability Analysis of Items Measuring Three Orientations</i>	147
Table 5: <i>Internal Validity Check of Calculated Versus Selected Orientations Using Chi-Squared Analysis</i>	151
Table 6: <i>Repeated Measures ANOVA Test on Scores of Five Orientations</i>	152
Table 7: <i>Repeated Measures ANOVA Test: Relationships between Orientations</i>	153
Table 8: <i>Repeated Measures ANOVA Test on Scores of Three Orientations</i>	154
Table 9: <i>Repeated Measures ANOVA Test: Relationships between Three Orientations</i>	155
Table 10: <i>Three-Factor Forced Principle Component Analysis with Direct Oblimin Rotation</i>	157
Table 11: <i>One-way ANOVA Test of Differences in Scores on Component One Based on Teachers' Calculated Orientations</i>	161
Table 12: <i>One-way ANOVA Test of Differences in Scores on Component Two Based on Teachers' Calculated Orientations</i>	162
Table 13: <i>Items Included in Two New Scales to Assess Curricular Orientation</i>	165
Table 14: <i>Calculated Orientation Frequencies Based on Five Orientations Framework</i>	178
Table 15: <i>Chi-Squared Test of Association between Calculated Orientation and Political Identification</i>	182
Table 16: <i>Racial Demographic Comparisons between Massachusetts and Nationwide HSS Teachers</i>	184
Table 17: <i>Chi-Squared Goodness of Fit Test between Massachusetts and Nationwide HSS Teachers in Terms of Race</i>	185
Table 18: <i>Gender Demographic Comparisons between Massachusetts and Nationwide HSS Teachers</i>	186
Table 19: <i>Chi-Squared Goodness of Fit Test between Massachusetts and Nationwide HSS Teachers in Terms of Gender</i>	187
Table 20: <i>Age Comparisons between Massachusetts and Nationwide HSS Teachers</i> ...	188

Table 21: <i>Chi-Squared Goodness of Fit Test between Massachusetts and Nationwide HSS Teachers in Terms of Age</i>	189
Table 22: <i>Political Ideology Comparisons between Massachusetts and Nationwide HSS Teachers</i>	190
Table 23: <i>Chi-Squared Goodness of Fit Test between Massachusetts and Nationwide HSS Teachers in Terms of Political Ideology</i>	191
Table 24: <i>Chi-Squared Goodness of Fit Test between HSS Teachers and MCAS-US Test Items</i>	250
Table 25: <i>Responses to Items Regarding the MCAS-US Test</i>	258
Table 26: <i>Pearson's Correlation between Teachers' Orientation Scale Scores and Beliefs About the MCAS-US Test</i>	263

Figures

Figure 1: <i>An Example of a Content Analysis Alignment Matrix from Porter (2006)</i>	138
Figure 2: <i>Field Map of Orientations</i>	169
Figure 3: <i>Field Map of Orientations with Massachusetts HSS Teachers (n =235) Overlaid</i>	198
Figure 4: <i>Four Depth of Knowledge Levels Derived from Webb</i>	211
Figure 5: <i>Five Content Area Categories Reflecting the Five Teaching Orientations</i>	212
Figure 6: <i>Frequency Distribution of Points across the Two Dimensions by Rater</i>	220
Figure 7: <i>Total Frequency Distribution of Points across All Four Panelists</i>	223
Figure 8: <i>Proportions of Test Points Awarded for Each Content and DOK Category</i> ..	224
Figure 9: <i>Proportion of Test Points Awarded for Each Depth of Knowledge (DOK) Category</i>	225
Figure 10: <i>Proportion of Test Points Awarded for Each Content Category</i>	226
Figure 11: <i>Comparison of Test Items' Content to Massachusetts HSS Teachers' Orientations</i>	248

Sample Test Items

Sample Test Item 1: <i>Item 3 from the MCAS Guide to History and Social Science Assessments (2007)</i>	205
Sample Test Item 2: <i>Item 27 from the High School US History 2007 Pilot Test Released Items</i>	217

Sample Test Item 3: <i>Item 1 from the MCAS Guide to History and Social Science Assessments (2007)</i>	227
Sample Test Item 4: <i>Item 11 from the MCAS Guide to History and Social Science Assessments (2007)</i>	227
Sample Test Item 5: <i>Item 22 from the High School US History 2007 Pilot Test Released Items</i>	228
Sample Test Item 6: <i>Item 8 from the MCAS Guide to History and Social Science Assessments (2007)</i>	229
Sample Test Item 7: <i>Item 16 from the High School US History 2007 Pilot Test Released Items</i>	230
Sample Test Item 8: <i>Item 2 from the MCAS Guide to History and Social Science Assessments (2007)</i>	231
Sample Test Item 9: <i>Item 1 from the MCAS Guide to History and Social Science Assessments (2007)</i>	236
Sample Test Item 10: <i>Item 10 from the MCAS Guide to History and Social Science Assessments (2007)</i>	240
Sample Test Item 11: <i>Item 5 from the MCAS Guide to History and Social Science Assessments (2007)</i>	278
Sample Test Item 12: <i>Item 6 from the MCAS Guide to History and Social Science Assessments (2007)</i>	279
Sample Test Item 13: <i>Item from Ravitch & Finn (1987), What our 17-Year-Old's Don't Know</i>	284
Sample Test Item 14: <i>Item 33 from the High School US History 2007 Pilot Test Released Items</i>	285
Sample Test Item 15: <i>Item 7 from the High School U.S. History Pilot Test Released Items</i>	287

CHAPTER ONE: BACKGROUND AND PROBLEM

A healthy democratic society requires citizens with both the knowledge to understand the problems it faces and the dispositions to actively solve them. Education of the people has long been a lofty goal, and a major undertaking of local, state and national governments. Without education, democracies can quickly degrade to an undesirable state. Indeed, the author of the Constitution, James Madison wrote "A popular government without popular information or the means of acquiring it is but a prologue to Farce or Tragedy or perhaps both... A people who mean to be their own Governors must arm themselves with the power knowledge gives" (Hunt, 1900 - 1910).

The power of knowledge and the disposition to act on knowledge however, are not natural traits in most individuals. Rather, these traits must be cultivated. If a society fails in this task, countless problems may arise. The United States currently faces such circumstances. This is borne out by many indicators, including trends as far-reaching as the national voting rate and as local as membership in neighborhood watch associations. Political leaders and policy makers usually expect public education to address the problem. While educational reforms targeted to primary and secondary school cannot change the participation rates of adults, at the very least these efforts can target up-and-coming generations and encourage them to engage in the political process. Yet, designing and implementing democratic education often unearths complex problems for history and social studies educators and those who assess their efforts.

In this chapter, I first address the problem of diminishing political and civic participation in the United States through an examination of social capital in American

society. I then turn to the consistently low performance of American students and adults on assessments of historical knowledge. Having demonstrated that two of the necessary characteristics of democratic citizenship – participation and knowledge – are sorely lacking among the American citizenry, I next question what can be done to remedy the problem. I include a discussion of the implications for education in general, and history and social studies education in particular.

The Problem of Diminishing Democratic Knowledge and Participation

Robert Putnam’s landmark study of civic and political participation in the United States has transformed the way social and political scientists discuss democratic participation. Putnam’s work focused on the concept of *social capital* – or a measure of the degree to which democratic characteristics, such as democratic social norms, social trust and networks of association, permeate a society (Putnam, 2000, 341).

Social capital both requires, and results in, desirable traits for democratic societies. First, social capital has a compounding effect – that is, the more one participates politically, the more likely one is to get involved in even more political activities. During his trip to the United States in the early 19th century, notable democratic theorist Alexis deTocqueville, a man whom Putnam referred to as “the patron saint of contemporary social capitalists” (Putnam, 2000, p. 292), asserted that big political issues that motivate large numbers of citizens also encourage individuals “to undertake less pressing social and civic interests in smaller numbers” (Tocqueville, 2000, 497). This tendency to participate in ever more political affairs is an effect of individuals’ learning the skills of participation (e.g. how to run meetings, speak in public,

write letters, organize projects and debate public issues with civility). Once they learn these skills, they are able to put them to use in other places more readily (Putnam, 2000, 339).

Social capital results in two additional desirable traits for democratic societies. Not only do individuals tend to participate in smaller numbers in “less pressing” social and civic interests, but high degrees of involvement in political and civic associations expose individuals to different members of their community, which fosters a trusting community. Tocqueville highlighted the effect of what he called “social power.” He noted that by joining community initiatives, individuals were exposed to “ideas or sentiments” (Tocqueville, 2000, 491) that, in isolation, they would not consider. This exposure encouraged the development of a tolerant, open and flexible society. Putnam echoed this position by arguing that citizens of states with high social capital were “far more committed to racial and gender equality” than citizens of low social capital states (Putnam, 2000, 356). Additionally, Putnam asserted that political and civic involvement increased awareness of humanity’s common fate by “widening awareness of the many ways in which our fates are linked” (p. 288). This increasing awareness creates a body politic that is “more tolerant, less cynical and more empathetic to the misfortunes of others” (p. 288).

Finally, social trust, which is closely linked to, and described as a subset of, social capital (Bryk & Schneider, 1996; Putnam, 2000), is also credited for many positive improvements in a society. The chief benefits of social trust are increased innovation and economic prosperity (Fukuyama in Bryk & Schneider, 1996; Johnston & Percy-Smith,

2003), increased institutional efficiency (Bryk & Schneider, 1996) and even increased physical health and mental well-being (Putnam, 2000). Increased association and its byproduct, increased trust, work together as a social lubricant, helping individuals to resolve problems collectively and allowing society to advance smoothly (Putnam, 2000). Political and civic association serves to extricate individuals from their solitude, encourage compassion, and promote further democratic participation; all of which advance a healthy and trustful society.

In recent years it has become increasingly clear that in the United States social capital, an ingredient on which democracies depend, has eroded substantially (Putnam, 2000). Civic and political activism has dropped considerably from the time Tocqueville penned *Democracy* in the early 19th century. It certainly has dropped measurably in the four decades since John F. Kennedy called on Americans to ask what they could do for their country. Putnam outlined a number of indicators of social capital that expose the alarming decline in civic and political participation in American society, including voting rates, volunteerism, and informal discussion of politics among friends. A society's voting rate is one of the most obvious indicators of its citizens' capacity to act democratically. While Americans' voting rate compares favorably with other democracies, Putnam claimed that, compared to our own past, Americans were not doing as well (Putnam, 2000). Putnam wrote:

In 1960, 62.8 percent of voting-age Americans went to the polls to choose between John F. Kennedy and Richard M. Nixon. In 1996, after decades of slippage, 48.9 percent of voting age Americans chose among Bill Clinton, Bob

Dole and Ross Perot, very nearly the lowest turnout in the twentieth century.

Participation in presidential elections has declined by roughly a quarter over the last thirty-six years. Turnout in off-year and local elections is down by roughly this same amount (p. 31-32).

Although the United States Census Bureau reports that voter turnout has increased slightly since 1996 (Bergman, 2005), turnout still remains disappointingly low relative to the number of citizens who are eligible to vote. In fact, while Barack Obama's candidacy for President was thought to reinvigorate voting rates in the 2008 general election, the year actually saw the lowest voter turnout rate in twelve years (Yen, 2009).

Additional indicators of social capital include rates of volunteerism in civic and political associations, and the amount of time individuals spend discussing politics with friends or acquaintances. Each of these indicators has decreased as well (Putnam, 2000). And, while societies that have high social capital enjoy benefits of that social capital such as increased trust in institutions, trust in one another and increased tolerance, the reverse is true in societies with low social capital. Instead of a society with growing trust and political activism, Americans are experiencing increasing isolation and cynicism (Putnam, 2000).

Declining social capital has created an undemocratic domino effect – that is, as social capital decreases, political association decreases. When political association decreases, the resultant social trust between citizens declines as well. When social trust declines, the desire to work on other citizen's behalf then declines leading to even lower social capital. This makes the problem of diminishing social capital difficult to correct.

Put simply, as Putnam (2000) wrote, “societies with low social capital are sick” (p. 288). What becomes of a democratic society that experiences a sickness such as this?

Putnam argued that first and foremost, individuals in a society with less social capital tend to trust less in the institutions that represent them. That is to say, ideally citizens in a democratic society feel as though the institutions that they control are representing them adequately and working on their behalf. Citizens in a society sick with diminishing social capital do not believe that these institutions are within their control or working on their behalf. To illustrate, Putnam (2000) explained that Americans in the 1960s were “strikingly confident in the benevolence and responsiveness of their political institutions...” (p. 47). He wrote:

Three in four said you could “trust the government in Washington to do what is right all or most of the time.” Such views nowadays seem antiquated or naïve. In the 1990s roughly three in four Americans *didn't* trust the government to do what is right most of the time (p. 47).

This increasing distrust begins with representative democratic institutions such as local, state and national government. Eventually, this distrust extends to individuals' relationships with one another.

Perhaps, then, it should come as no surprise that what Putnam (2000) referred to as “booming social distrust” (p. 141) has come to characterize American civil society and that Americans in fewer and fewer numbers see their compatriots as “honest and moral” (p. 139). This distrust wreaks havoc on a citizen's day to day exchanges with other individuals. Americans report that they experience more friction in their relationships

with one another than they once did (p. 140). Putnam suggests that indicators of this increasing distrust are observable in everyday exchanges from things as mundane as increasing road rage and the disappearance of hitchhiking to larger social trends such as the “massive expansion of the legal profession [and] preventative lawyering” (p. 146). Abruptly, Putnam claimed, Americans “began to demand to ‘get it in writing’” (p. 146 – 7). In the past, many Americans trusted community watch associations and the common decency and civility of fellow citizens to protect and monitor their neighborhoods. As Putnam pointed out, cynical Americans now look to law enforcement and other state agencies (often not to be trusted either) to maintain order between civilians.

This increase in cynicism and distrust has led many Americans to take up more extremist political stances – another factor that can be debilitating for a democracy. Putnam suggested that the adoption of extreme political positions and the virtual disappearance of moderates from the political sphere is the direct result of the dissolution of political associations and the divorce of people from the larger community. Without the political and civic associations that allow for people from opposing political camps to deliberate, it becomes easy for people to “demonize anyone who disagrees” (p. 342). When people become armchair spectators of the American political drama, they rarely have an opportunity to exchange words and ideas with those whom they disagree and “anonymity,” Putnam claimed, “is fundamentally anathema to deliberation” (p. 342). With moderates deserting the political sphere and those from opposite ends of the political field refusing to associate and deliberate over their disagreements, Putnam

concluded that Americans are failing to live up to the type of democratic ideal that Tocqueville observed and admired in the early 19th century.

As stated above, democracies require citizens who possess both the knowledge to understand problems and the dispositions to actively solve them. The above discussion shows that citizens in the United States seem to be losing the democratic habits required to actively solve social problems. Conceivably citizens may simply lack the knowledge necessary to act confidently and competently in the public arena. Putnam (2000) suggested, “knowledge about public affairs and practice in everyday civic skills are prerequisites for effective participation” (p. 403). Questions about whether or not Americans have sufficient knowledge to act on their beliefs must be considered. While it is difficult to establish what type of knowledge a citizen should possess to understand contemporary problems, democratic knowledge is often defined as knowledge of history and contemporary political affairs. In this regard, American citizens may be worse off than they are in their participation habits.

Late night television comedian, Jay Leno, has regularly made light of the alarming lack of basic historical knowledge of American citizens. In his popular skit, “Jay Walking,” Leno walks the streets of New York City and challenges unsuspecting individuals with basic history questions such as, “Name the first president.” This skit does well in late night comedy because of the wildly incorrect answers the participants offer. But, standardized assessments in United States history including the National Assessment of Educational Progress (NAEP) bear out the fact that these embarrassingly off-the-mark responses may be more typical of the historical knowledge of most

American high school graduates than one would wish (Gaudelli, 2002; Leming, Ellington, & Porter-Magee, 2003; Ravitch, 2001; Ravitch & Finn, 1987; Shenkman, 2008).

The most recent administration of the National Assessment of Educational Progress (NAEP) in U.S. history (2006) showed that American students were not doing much better in U.S. history than they did on the 2001 administration of the test, with only 46% of American students scoring at or above the “basic” level compared to 43% in 2001 (Lee & Weiss, 2007). The assertion that American students are failing on history assessments is not new. An historical analysis of American students’ performance on history assessments as far back as 1917 and including the 1987, 1994, 2001 administrations of the NAEP in U.S. history demonstrated that American students consistently perform poorly on these tests (Wineburg, 2004).

In addition, a recent report released by *Common Core* stated that 17 year-olds earned a cumulative grade of “D” on a test of historical literacy (Hess, 2008). Based on a random sample of American teens, *Common Core* found that 20 percent could not identify America’s enemies in World War II, and more than 25 percent mistakenly believed that Columbus sailed to America after 1750. Just two in five could place the Civil War in the correct 50-year period and nearly a quarter could not correctly identify Adolf Hitler (Hess, 2008). The historical themes and concepts that the *Common Core* assessment tested are topics that should be highly prioritized in the history and social studies curriculum. Certainly, the Civil War, World War Two, Christopher Columbus and Adolf Hitler figure prominently and explicitly in the Massachusetts history and social

studies frameworks (*Massachusetts History and Social Science Curriculum Framework*, 2003). So, while many historians and history educators argue that the historical knowledge tapped for standardized assessments is often so random as to be meaningless (Rothstein, 2004b; Wineburg, 2004), one would be hard pressed to argue that familiarity with a monumental historical figure such as Adolph Hitler is insignificant knowledge for democratic discourse.

Finally, recent survey results released by the Pew research center highlighted the fact that Americans fare little better in their knowledge of national politics and current events. In the period since 1989, the percentage of Americans who could correctly identify the Vice President of the United States decreased from 74 to 69% and those who could identify their own state governor decreased from 74 to 66% (Kohut, 2007). Furthermore, less than half of those polled could correctly identify key figures in current events including speaker of the house, Nancy Pelosi, or presidential candidate Barack Obama just months before he won the democratic nomination for president (Kohut, 2007).

If healthy democratic societies require citizens who possess both political knowledge and habits amenable to political participation, these large scale studies of the voting and participation behaviors, and the political knowledge base of average American citizens is worrisome to say the least. How can Americans, who are concerned about the civic health of the nation, act to remedy the growing problems of social alienation, civic disengagement, and historical ignorance and turn this trend around?

Addressing the Problem: A Call for Improved Civic Education

Americans concerned about declining civic knowledge and participation focus on solving the problem through effective public policy. Specifically, political and social scientists and policy makers alike often turn to educational policy initiatives to reverse the tide of civic ignorance and apathy. Indeed, Tocqueville (2000) argued that “if political association is an art, let the art be taught” (p. 503). Public schools have taken this charge seriously. Education for active citizenship is at the very root of their mandate. In fact, as *Education Next* writer David Campbell asserted, “producing better citizens was the original justification for creating American public schools” (Campbell, 2001). Graduation requirements that compel students to do community service, curriculum reforms that include social histories of underrepresented minorities and educational aims that encourage students to perfect their analytical and debate skills each serve to introduce youth to alternative ideas and lead them to invest in the art of civic participation. Learning this “art” of civic participation is central to a healthy democratic society and, as Tocqueville declared, “...we must enlighten society at any cost because the art is fleeing from us” (Tocqueville, 2000, 503).

While many school leaders and educators across the disciplines have taken up the charge to teach “the art” of political association, history and social studies (HSS) educators specifically are called upon to spearhead these efforts. History and social studies professionals themselves generally see citizenship education as the rationale of their field. Yet, there is a common perception that history and social studies teachers are failing to prepare citizens. Critics point to students’ consistently poor performance on the

aforementioned widely-publicized standardized history tests as evidence of the fact that HSS teachers are failing their students. If all HSS teachers are taking their charge to prepare informed and participatory citizens seriously, they contend, why do these embarrassingly low scores on tests persist?

A current debate in the field of HSS education centers on the suggestion that teaching orientations may have something to do with the continued “abysmal” performances of students on history assessments (Manzo, 2002). Critics argue that American students’ failure to demonstrate civic knowledge and participation is a direct result of the fact that some teachers are simply less effective than others at preparing citizens. These critics often claim that teachers’ orientations add to, or even create the problem of historic illiteracy rather than solve the problem (Kauffman, 2002; Leming, et al., 2003; Ravitch, 2001; Stotsky, 2004). Critics in the education field often blame HSS teachers for these poor results and claim that the problem is exacerbated when children are taught skills to “construct history” rather than taught “historical truth” (Barth, 1997; Kauffman, 2002; Leming, et al., 2003; Ravitch, 2001; Stotsky, 2004). Diane Ravitch (2000), for example, highlighted the distinction between teaching orientations by noting that teachers who teach from a constructivist orientation believe they must “never lecture or ‘tell’ [and] that any memorization [is] intolerable” (p. 442). Leming & Ellington (2003) blamed the “trouble” in the HSS field on the “hostility on the part of many educators at all levels to the kinds of basic knowledge ordinary Americans think important for children to learn” (p. ii). Stated simply, critics contend that students’ failure to perform on HSS assessments has to do with HSS teachers’ orientations – specifically

HSS teachers' reluctance to transmit basic historical facts. The argument here is that what history and social studies teachers *believe* about social studies, and what they subsequently *do* in their classroom, matters considerably to the outcome of civic education.

Based on a review of related literature, professionals within the field of HSS education have examined additional indicators of disappointing outcomes of history and social studies education. Marker and Mehlinger (1992) argued that social studies education was failing to increase voter participation and respect for the law, had no discernable impact on the socio-moral development of youth, and was seen by students as less important and less interesting than other subjects (p. 845). Citing the same Finn and Ravitch study discussed previously, Marker and Mehlinger noted that Americans “cannot locate major countries on a map; locate historical events in time; [or] name current political leaders” (p. 844). Notable education researcher, Stephen Thornton (1991) also questioned the outcome of HSS education citing research that showed that “after two or three exposures to U.S. history, students still remain indifferent to and ill informed about it” (p. 236). Each of these scholars call into question HSS teachers' ability to successfully meet any of the goals they may have for their pupils, – as varied as those goals may be.

It should be noted that no studies were found in the research for this study that directly linked HSS teachers' approaches to the subject matter and student outcomes in historic literacyⁱ (Thornton, 1991, p. 236). That is to say, there is no empirical evidence to suggest that students who are educated by a teacher who works, for example, from a constructivist perspective, know less history when they leave that teacher's class than

students who are educated by a teacher who teaches from a different orientation. What little evidence exists demonstrates that students perform equally on HSS assessments whether their teachers' claimed to have followed state standards or not (Rothstein, 2004). Some researchers, therefore, have claimed that students' failure on the standardized tests does not signal a problem with teaching approaches. Rather, the argument is that standardized history tests such as the NAEP, not teachers' constructivist approaches, are to blame. Critics have asserted that these standardized assessments in history are a seriously flawed measure of a student's capacity for citizenship and that test developers have yet to create a good measure (e.g., Rothstein, 2004a, 2004b; Wineburg, 2004).

Even so, the critics' construction of the civic participation and knowledge problems that are the result of "soft" teaching practices has held considerable sway in public discourse over the last several decades (Bennett, Finn Jr, & Cribb Jr, 1990; Leming, et al., 2003; Ravitch & Finn, 1987; Salmans, 1987). Those who construct the problem in this manner make assumptions about how history and social studies is taught, and subsequently how students learn. One assumption they make is that if teachers teach in traditional ways (e.g. lecture and drill) students will absorb the information well enough to succeed on tests of civic knowledge. A second assumption is that learning more historical content will lead students to participate in civic life. Yet with no evidence to support the contention that teachers' constructivist stances are to blame other than students' low test scores on history assessments, questions must be raised about whether it is fair to assume constructivist teaching and low history test scores go hand in hand. The inverse of this argument is that teachers are not to blame for poor performance on

tests. Rather, standardized tests are to blame for measuring constructs that teachers do not teach.

Regardless of the dearth of evidence to support these claims, the critique of HSS teaching underscores the belief of many critics that teaching historical facts and encouraging memorization results in historical literacy and is the best approach to teaching HSS. Concurrently, constructivist orientations to teaching history and social studies result in either historical illiteracy or some other outcome that is impossible to measure on a standardized test, and therefore, undesirable.

Massachusetts: A State on the Brink of Civic Education Reform

This discussion emphasizes that there are many enduring issues in the field of history and social studies, issues that are considered in this study. In the most general sense and at the highest level, declining participation rates of Americans in civic activities creates a real problem for the health of a democratic society. Many look to public education in general, and history and social studies education in particular, to address or at least stem the tide of this growing problem. Yet, after decades of reform initiatives, tests of historical literacy consistently reveal a lack of civic literacy. Is the problem, like the critics contend, with HSS teachers' approaches to teaching? Could the problem be with the tests devised to assess civic knowledge? Or, is the problem perhaps something that no one has previously considered?

While history and social studies standards have been around for a long time, the introduction of high stakes tests in history and the social sciences in Massachusetts makes the state a strategic site for learning if and how education policy affects teachers'

practice, and how teachers' practices affect education policy. In Massachusetts, "defining the social studies" has become less an academic and theoretical problem and more of a pressing real-world issue. Do individual HSS teachers, who come from widely disparate curricular orientations within the field, emphasize uniform concepts regarding United States history in the classroom? Are the concepts that the test assesses similar to the constructs emphasized by teachers? To analyze the relationship between Massachusetts HSS teachers' orientations and the state standardized test in United States history, this study will explore the following questions:

1. What are the orientations of history and social studies teachers in the state of Massachusetts?
2. What skills and constructs are measured by the proposed high-stakes, state-mandated high school test in United States history?
3. What is the relationship between the orientations of history and social studies teachers in the state of Massachusetts to the constructs measured by the Massachusetts state-mandated history and social studies assessment?

Organization of the Dissertation

This chapter illustrates the enduring nature of both poor civic knowledge and low levels of civic engagement and lays the groundwork for exploring the relationship between the teaching orientations of Massachusetts HSS teachers and the skills and constructs measured by the proposed high-stakes test in United States history.

In chapter two, I provide the historical and scholarly context for exploring this relationship by first clarifying the theoretical framework of the study and then reviewing

two bodies of literature that are related to the questions of the study. First, I present the theoretical framework that guided the dissertation, which draws on the work of both democratic theorists in the field of political philosophy and the philosophical field of pragmatism. I tapped these two theories to form a singular theoretical construct, which Smiley (1999) has referred to as “democratic pragmatism,” because both theories say a great deal about the role of public education in a democratic society (MacGilvray, 1999; Dewey, 1916, Gutmann, 1987). Democratic pragmatism as a set of concepts, theories and assumptions prompted the questions I asked in this study and is one of a few bodies of work where scholars have theorized answers to those questions. By exploring the relationship between HSS teachers’ orientations and standardized history tests through this lens, one assumption is clear: While different approaches to history and social studies are expected, only certain types of teaching are likely to result in the outcome democratic pragmatists seek. Democratic pragmatism, therefore, is used to assess the quality of public schooling for democratic citizenship (Smiley, 1999). The theory’s explicit connection the aims and ideals of American society make this particular theoretical frame highly applicable in this study.

Following the description of the theoretical frame, chapter two reviews two separate bodies of literature to contextualize and situate the dissertation. In the first review of the literature, I examine related conceptual and empirical scholarship on teachers’ curriculum orientations in history and social studies. Using the work of Barr, Barth and Shermis (1977) as a starting point, I organize the scholarship on teaching orientations into five categories: cultural transmission (CT); social studies as social

science (SS); reflective inquiry (RI); reflective inquiry for personal ends (RIPE); and reflective inquiry for social ends (RISE). These categories provide the conceptual frame for the survey of Massachusetts history and social studies teachers and the content analysis of the MCAS test. Included in the first review of the literature are findings from quantitative studies that describe the prevalence of orientations within the nationwide HSS teaching population, as well as findings from qualitative studies describing the classroom work of HSS teachers from various curricular orientations.

The literature on teachers' curricular orientations is followed by a review of conceptual and empirical research on standards and assessments in history and social studies. Madaus' (1988) landmark article on the seven principles of high stakes testing was used to organize the conceptual work. The review of empirical studies includes discussion of the call for standards and assessments, as well as analyses of the effects of standards and assessments on teachers' practice. Interestingly, studies of teachers' beliefs about the anticipated effects of high stakes testing in history and social studies contrasts significantly with qualitative studies that describe HSS teachers' actual classroom practice under these reforms. While the studies of teachers' beliefs demonstrate that HSS teachers suppose high stakes tests are drastically affecting education, qualitative studies that describe practices of HSS teachers show mixed effects of these tests – that is to say, some teachers' practice is significantly influenced by high stakes tests while other teachers don't appear to alter their practice at all.

Chapter three describes the two research designs for this dissertation. As described in detail in this chapter, the first research design, a systematic survey of

Massachusetts history and social studies teachers, serves to answer the question, “What are the orientations of history and social studies teachers in the state of Massachusetts?” Included in the description of the first research design is an evaluation of previous surveys of HSS teachers, modifications made to previous surveys, piloting procedures and results, and sampling, survey administration and analysis procedures for my dissertation. The second research design is a content analysis of the Massachusetts Comprehensive Assessment System’s 10th/11th grade United States history test (MCAS-US test), and serves to answer the question, “What skills and constructs are measured by the proposed high-stakes, state-mandated high school test in United States history?” Here, I use the work of Porter and colleagues (Porter, 2006; Porter & Polikoff, 2008; Porter & Smithson, 2001) who conduct content analyses of high-stakes, standardized tests to assess the degree of alignment between intended curriculum (standards), enacted curriculum (classroom practice), and assessed curriculum (standardized tests).

Chapter four and five present the findings of this study. In chapter four, I explore the orientations of history and social studies teachers in Massachusetts and I make two key arguments. First, I argue that in order to understand HSS curricular orientations, a more nuanced interpretation of teachers’ stances within those orientations, which encompasses but goes beyond the three traditions thesis, is needed. I propose an alternative framework for understanding HSS teachers’ curricular orientation, which I call the “Field Map of Orientations,” and I analyze the results of the survey based on this framework. Second, I argue that teachers in Massachusetts are unlike HSS teachers nationwide in that they promote a relatively consistent message about the purpose,

pedagogy and content of HSS, and in that they share an overwhelming affinity for the reflective inquiry teaching orientation.

In chapter four, I also use the theoretical lens of democratic pragmatism to ground the first of three comparisons. Here, I explore how the pedagogical and curricular visions of teachers who work within the reflective inquiry orientation compare to the ideals of teaching history and social studies outlined and promoted by educationists who work from a democratic pragmatist theory. Findings from this comparison indicate that the curricular beliefs of teachers who work within the reflective inquiry orientation are very much in line with the ideals of teaching history and social studies promoted by theorists who work within the democratic pragmatist frame.

In chapter five, I analyze the skills and constructs measured by the MCAS-US test based on the content analysis. I argue that the MCAS-US test overwhelmingly assesses content associated with the cultural transmission orientation. Furthermore, the test is comprised of items that only require factual recall or basic reasoning skills from students, rather than complex or extended reasoning. In this chapter, I make the second of three major comparisons in this study by comparing the content message of the test with the educational tenets of democratic pragmatism. Based on this comparison, I argue that the MCAS-US test employs assessment techniques that work in cross-purposes to the suggestions outlined by theorists who work within the democratic pragmatist tradition.

In the final section of the chapter, I make the last of three major comparisons between Massachusetts HSS teachers' orientations toward the subject and the skills and constructs measured by the MCAS-US test. I argue that there is a very clear

misalignment between orientations of history and social studies teachers in the state of Massachusetts, on the one hand, and the skills and constructs measured by the MCAS-US, on the other.

Both teachers and test developers purport to be working from the history and social studies state frameworks, yet those frameworks are so broadly written that test developers and teachers can derive contrary meanings from them in their work. Yet, because a consistent theoretical approach appears to be guiding the work of HSS teachers in Massachusetts (as demonstrated by the findings from the survey), and because that approach bears significant resemblance to approaches advocated by theorists who work from a democratic pragmatist lens, I suggest in chapter six that changes need to be made to the way that test developers determine what items should appear on the high stakes test. I propose a set of four criteria, rooted in democratic pragmatist theory, to guide test developers in that process. In light of diminishing social capital, democratic participation and civic knowledge, the conversation about how to reform history and social studies teaching and assessment is a significant one. Creating an assessment that accurately reflects the goals of democratic and civic education, and ensuring that the nation's HSS teachers and test developers are working in a synchronized fashion to meet the goals of democratic education is a worthwhile goal. This study is an effort to enlighten the process by suggesting ways in which test developers and HSS teachers can work in greater unison toward this shared goal.

CHAPTER TWO: FRAMING THE STUDY

In the previous chapter, I wrote that many critics of HSS education argue that social studies educators fail to teach basic concepts to their students. Many critics blame the nation's school teachers for teaching American students "soft social studies" rather than "true history." These critics believe failure on standardized tests signals the failure of American public schools to ready students for a democratic participation (Gaudelli, 2002; F. Hess, 2008; Ravitch, 2001; Ravitch & Finn, 1987). For others, standardized tests results tell us little – neither about students' preparedness for democracy, nor about teachers' ability to teach the skills and content of the subject. One academic distinguishes these two groups by noting that their responses to students' performance on tests will vary depending on a person's "deeply held curricular assumptions...[including assumptions about] what we shall teach, how we shall teach it and who should decide" (Gaudelli, 2002).

In light of that, in the first large section of this chapter, I address the theories and assumptions underlying the study – in other words, can it be assumed that public educators should take civic education as the central mandate of public schooling? In the next large section, I review the literature on history and social studies teachers' orientations in the field. The review helps to ground the study by developing a framework for the vastly divergent philosophical, pedagogical and curricular orientations of history and social studies teachers. In the third and final section, I review the literature on state standards and assessments in history and social studies. Specifically, what does the literature say about which states employ tests in history and social studies, what

constructs do the tests measure (e.g. does the test measure students' acquisition of citizenship skills?), how do test-makers measure those constructs, and how do teachers respond to those tests?

The Need for a Theory: Democratic Pragmatism

The argument that public schools fail to prepare students for citizenship in a democratic society is based on the assumption that the purpose of public schooling in general is to prepare students for citizenship – rather than, say, to prepare students for technical work. To make this assumption about the proper business of public schooling, I first establish an explicit theory of democratic education to guide the discussion.

A theory of education acts as a set of principles from which to base policy decisions and to analyze those decisions as they are put into practice (Gutmann, 1987). To establish a vision of education that is consistent with the values and beliefs of the American people, I begin with a theoretical framework that dependably represents the American ethic of democracy. Developing a *democratic* theory of education ensures that the theory is in line with the social, political and cultural values associated with democratic societies.

To establish a democratic theory of education, I draw on the work of both democratic theorists in the field of political philosophy (including, primarily, the work of Amy Gutmann) and the philosophical field of pragmatism (including the work of John Dewey, among others). I chose democratic theory and pragmatism to form a democratic theory of education because both theories deal specifically with, and say a great deal

about, the role of public education in a democratic society (see MacGilvray, 1999; Dewey, 1916; Gutmann, 1987).

Pragmatism, at its core, is a set of epistemological beliefs that address questions about the production and justification of knowledge (MacGilvray, 1999, 545). Centrally concerned with an individual's ability to use inquiry and logic to formulate knowledge about the world, pragmatists have been criticized for being too atomistic (Talissee, 2004). Put simply, critics claim that pragmatists say little about how individuals function in a community. What pragmatists need, Talisse argued, is a way to "reunite the atomic individuals that reside at the basis of their theory, to socialize the essentially asocial" (Talissee, 2004, 1). While pragmatic inquiry is often understood as a way for an individual to obtain knowledge, pragmatism can also serve a role in achieving democracy (Smiley, 1999). To do so, pragmatists must reach beyond the confines of pragmatic theory, as Dewey did, in search of an auxiliary theory that will help explain how the individual pragmatic inquirer becomes animated in the public sphere.

Democratic theorists make up for this shortcoming by extending the pragmatic epistemological tradition to the public sphere such that, as one theorist contended, the "tradition of pragmatic moral thought [becomes] principled advocacy for liberal democratic ideals" (MacGilvray, 1999, 542). While democratic theorists do not usually describe themselves as pragmatists per se, many of the assumptions and arguments democratic theorists make show their pragmatic leanings (Festenstein, 2004; Knight & Johnson, 1996).

For example, Marion Smiley, a political philosopher who bridges pragmatic theory and democratic politics, acknowledges that humans are not simply individuals conducting inquiry and acquiring knowledge in a cultural vacuum. Rather, humans are ...a community of inquirers who symbolically interpret the results of inquiry through collective symbols shared within what Dewey calls his ‘public.’ Hence, the results of inquiry must be understood as mediated by the structure of community, e.g., by what the community values and the particular symbols that it uses... (Smiley, 1999, 631).

What Smiley referred to as the “structure of community” is, in the case of the United States, democratic. Democratic structures mediate activity on the public stage in specific ways and democratic ideals require that certain thresholds be met in the public arena.

One way that democratic activity is mediated is by the norms of democratic deliberation (Festenstein, 2004; Gutmann, 1987; Gutmann & Thompson, 1996). Gutmann wrote that the core idea behind deliberative democracy is simple: “when citizens or their representatives disagree morally, they should continue to reason together to reach mutually acceptable decisions” (Gutmann & Thompson, 1996, 1). Gutmann further explained this deliberative process by noting that the “moral authority of the collective judgment [arrived at through the deliberative process] depends in part on the moral quality of the process by which citizens collectively reach those judgments” (p. 4). Gutmann outlined three *conditions* of democratic deliberation including reciprocity, publicity and accountability. She also described three *principles* that govern democratic deliberation including liberty, basic opportunity and fair opportunity (p. 4).

The governing principles and conditions of deliberative democratic exchange require citizens to hold particular values and wield certain skills. Thus the deliberative democratic norm has particular repercussions for the aims of democratic education. Examples of how education should encourage values and skills that, in turn, reinforce democratic deliberation are addressed throughout this dissertation. Meanwhile, I pause here to emphasize the point of the preceding paragraphs: Democratic theory and pragmatism can be viewed as two sides of the same coin. On one side, pragmatism provides a theory of knowledge. On the other side, democratic theory provides a theory of knowledge-in-action on the public stage. Borrowing from Smiley, I refer to this democratic theory of education as “democratic pragmatism” (Smiley, 1999).

Democratic pragmatism served as the theoretical frame through which I constructed the problem, collected the data and analyzed the results of this study. Democratic pragmatism raises several questions that are central to this study such as: What is the purpose of schooling in a democratic society? What characteristics should effective citizens in a democracy exhibit? How should we teach students so they may embody these traits? (This includes questions about the purpose of curriculum, the method of instruction, and the content knowledge including epistemological questions of truth, fact and justification of both.) How do we evaluate whether or not teaching efforts have been successful? Finally, who should make decisions about public education and about educating citizens for democratic participation? It must be acknowledged that pragmatism and democratic theory answer a multitude of questions that are not raised in

this study. I limit the discussion of these theoretical camps to works that specifically and explicitly address the purpose of, the practice of, and policy avenues for public education.

Democratic Pragmatism and the Role of Public Schools

Proponents of democratic government often draw attention to the fundamental role that public schooling should play in a healthy democratic society. Thomas Jefferson claimed he knew of “no safe depository of the ultimate powers of the society but the people themselves” (Ford, 1899). Yet, he qualified this faith in people by noting that some people may be too unenlightened to wield this power appropriately. He wrote that the remedy to this ignorance was not to take power from them. Rather, he wrote, democrats must “inform their discretion by education. This,” he claimed, “is the true corrective of abuses of constitutional power” (Ford, 1899).

Early democratic theorists such as Alexis de Tocqueville, believed not only that education could enlighten the masses and thus make democracy function more efficiently, but that democratic education was a fundamental aim and outcome of education. Tocqueville mused, “One cannot doubt that in the United States the instruction of the people serves powerfully to maintain a democratic republic... the sum of men’s [sic] education is directed toward politics” (Tocqueville, 2000, 291). Democratic theorists and public education advocates have, from the very beginnings of American society, promoted citizenship training as a central aim of schooling. While one cannot claim that political education is the only desirable aim of public education, in an open and democratic society, political education must be a top priority.

It is clear that education should be devoted to democratic ideals, but the relationship is a reciprocal one: Education should have democratic aims, and democracies rely upon education. Because of this reciprocal and symbiotic relationship, Dewey explained, democratic societies are more centrally concerned with education than other types of societies. Dewey (1916) wrote,

The realization of a form of social life in which interests are mutually interpenetrating, and where progress is an important consideration, makes a democratic community more interested than other communities have cause to be in deliberate and systematic education. The devotion of democracy to education is a familiar fact (p. 87).

Contemporary democratic theorists, such as Amy Gutmann, take it as a given that a central aim of public education in a democratic society is prepare youth for citizenship and that the preparation of such citizens is central to healthy democracies. Gutmann established that the purpose of primary educationⁱⁱ in a democracy is to inculcate character and moral reasoning for democratic ends. Gutmann (1987) qualified,

Although inculcating character and teaching moral reasoning by no means exhaust the purposes of primary education... together they constitute its core political purpose: the development of 'deliberative,' or what I shall interchangeably call 'democratic,' character... In practice, the development of deliberative character is essential to realizing the ideal of a democratically sovereign society (p. 51 – 2).

Democratic pragmatism grounds the over-arching assumption of this study that the proper business of public schooling in a democratic society should be to prepare citizens. Democratic pragmatism, as a set of ideas, concepts and constructs, helps to define what democratic schooling should look like and how to work to realize those goals. Democratic pragmatism also raises questions as to whether those goals are being met adequately.

Democratic Pragmatism on What Democratic Education Should Look Like

Democratic pragmatists contend that creating curriculum and instruction in line with democratic principles will better prepare students for citizenship and remedy some of the failures of public schooling described at the outset of this paper. As noted above, Gutmann contends that democratic deliberation requires particular conditions (reciprocity, publicity and accountability) and principles (liberty, basic opportunity and fair opportunity) and that a primary aim of democratic education is to train youth to embody deliberative character (Gutmann, 1987). But, what should education that engenders a democratic, deliberative character look like?

Democratic theorists and pragmatists assert that there is room for both conservative and liberal ideals in democratic institutions (Dewey, 1916; Festenstein, 2004; Hickman & Alexander, 1998; Robinson & Groves, 2004; Rorty, 1999) Schools, being institutions in a democratic society, should ideally be no exception to this contention. On the one hand, public education in a democratic society must establish cultural cohesion so many diverse citizens can get along with, and understand, one another. Democratic pragmatists often write of the schools' role in bringing about greater

social cohesion and how that cohesion, in turn, creates democratic efficiency (Gutmann, 1987; Hickman & Alexander, 1998; Rorty, 1999). Creating an ease of communication across diverse populations requires that youth must be introduced to, and practiced in, the language, norms and values of American life. Dewey often drew attention to this conservative leaning of education. Dewey (1916) recognized that “education proceeds ultimately from the patterns furnished by institutions, customs, and laws” (p. 89). The young, he believed, depended on these previously existing social “arrangements” for their education. Ultimately, these conservative conceptions would be the basis from which individuals would work to advance society in a more progressive direction.

Richard Rorty (1999), a prominent neo-pragmatist, believed that socializing youth to conservative notions of the world “as it is” is the ultimate responsibility of primary education. Amy Gutmann (1987) also acknowledged the conservative tendencies of education (particularly primary education) and noted that

Training of [a] ‘didactic’ sort is democratically desirable because it enables citizens to understand, to communicate, and in some cases to resolve their disagreements. Without this sort of mutual understanding, we could not expect to achieve widespread toleration of dissent and respect for differing ways of life (p. 50).

Education advocates often overlook the conservative aim of schooling. Yet, the conservative objective of schooling is not to reproduce society as it is. Rather, it serves the liberal objective – it is a means to a more progressive end. Dewey explains how the conservative objective of schooling serves the liberal objectives.

Dewey warned that the conservative purpose of education is not the ultimate aim of democratic education. He wrote, “Each generation is inclined to educate its young so as to get along in the present world instead of with a view to the proper end of education: the promotion of the best possible realization of humanity as humanity (Dewey, 1916, 95). Like Dewey, democratic pragmatists contend that education in a democratic society must liberate the minds of individuals (Rorty, 1999). To do so, individuals must be exposed to multiple experiences and interpretations of the “good life” (Gutman, 1982; Rorty, 1999). Some neo-pragmatists call into question whether or not young children have reached a stage of development such that they are capable of deep and meaningful critical thought. They argue that this liberal objective be left to secondary (taken to mean collegiate) education (Rorty, 1999). Leaving aside the question as to precisely *when* liberal and critical education should begin, democratic pragmatists argue that this liberal component of schooling – the process of teaching young democrats “how to think logically, to argue coherently and fairly, and to consider the relevant alternatives before coming to conclusions” – is paramount in truly democratic education (Gutmann, 1987, 50).

Ultimately, however, for a democratic pragmatist, the most important aim of education in a democratic society is to develop the intellectual and emotional understanding in individuals to vote properly, to care about fellow citizens and to make progress toward a better world (Dewey, 1944; Gutmann, 1987; Rorty, 1999). This requires both a liberal and conservative leaning in education. Democratic pragmatists assert that citizens must understand and identify with one another to keep alive the social

trust that maintains the social contract. Meanwhile, individuals must also be liberated enough from the current status-quo in order to envision and work towards a society that reflects better our lofty ideals (Dewey, 1944; Gutmann, 1987; Rorty, 1999). From the democratic pragmatist perspective, if we make “fully realized” individuals our goal in education, a healthy democratic state will follow (Dewey, 1916; Hickman & Alexander, 1998; Martínez Alemán, 2001).

Democratic pragmatists clearly outline a theory of democratic education. They acknowledge that democratic education requires both conservative and liberal objectives with the ultimate aim of instructing youth to deliberate democratically so that Americans can work toward a just society. Using democratic pragmatic theory as a foundation to advocate for particular educational content and practices becomes, for lack of a better phrase, a thorny matter. Educational content and educational practices are discussed in detail in chapter four and five.

Concluding Thoughts: How Democratic Pragmatism will inform this study

Researchers in the field of history and social studies education often claim that research conducted in the field is “atheoretical” (Fraenkel & Wallen, 1991; Popkewitz & St. Maurice, 1991). While a great deal of empirical research has been conducted in this field as to what to teach, how to teach it and who should decide, because the work is not grounded by common assumptions about the purpose and aims of history and social studies education, claims as to “how we’re doing” in social studies are difficult to substantiate. By establishing and working from an explicit democratic pragmatic framework, I seek to avoid this problem.

Democratic pragmatism as a set of concepts, theories and assumptions about education prompts the kinds of questions I ask in this study. Namely, what democratic competencies are to be developed that result in effective citizens? How do teachers claim to meet this aim? How should these outcomes be assessed both by teachers and by larger governing agencies such as the state? Democratic pragmatism is one of a few bodies of work where scholars have theorized answers to the questions I pose. The theory's explicit connection the aims and ideals of American society make this particular theoretical frame highly applicable in this study.

This study examines the relationship between the aims of teachers of history and social studies and the state standards and assessments in history and social studies. Yet, the democratic pragmatist arguments about the purposes of public education raises additional, deeper questions about the adequacy of both teachers' aims and methods in educating citizens as well as the aims of state standards and assessments in a subject that purportedly is meant to prepare citizens. The upcoming high-stakes, statewide U.S. history test for 10th and 11th graders in Massachusetts functions as a type of nexus where these deeper questions collide.

Democratic pragmatists note that dilemmas and decisions in a democracy should be solved through democratic disagreement and deliberation (Gutmann, 1987). Leaving the decision to an "authority" is wholly undemocratic and contrary to democratic theorists' larger goals (Gutmann, 1987). Democratic theorists contend that education decisions should involve multiple "stakeholders" including the family, the state and professionals in the field (Gutmann, 1987). All three bodies of stakeholders have shared

authority in decision making in democratic public schooling. Democratic pragmatism, then, as a theoretical outlook becomes both a method of resolving conflict about how best to educate citizens and a measure by which we can make decisions about the “consequences of our actions” (Smiley, 1999, 631).

By exploring this relationship through the lens of democratic pragmatism, a few assumptions are clear. First, different approaches to the study of history and social studies are to be expected. As with any moral problem in a democracy, individual teachers legitimately have their own solutions as to how best to educate citizens for democratic participation. Yet, only certain types of teaching are likely to result in the outcome that democratic pragmatists seek. Democratic pragmatism may be used as a mode of analysis to discern the degree of quality in public schooling for democratic citizenship (i.e. is the content, the pedagogy and the assessment individually and as a whole sufficiently justifiable as democratic in nature?) (Smiley, 1999).

Democratic pragmatism can also be used as a benchmark to analyze the process and product of education policymaking. Do state standards and assessments in history and social studies in the state of Massachusetts support the formation of the skills of inquiry and democratic deliberation? The assumption embedded in pragmatic theory is that historical concepts are not historical truths. Rather, historical concepts are repeated narratives over generations. By taking a pragmatic theoretical stance when looking at large scale assessments in history, I assume that history assessments can be done correctly (i.e. there are historical concepts that merit transmission to the next generation), but those concepts will change over time depending on their utility for citizens. This

assumption will ground the study epistemologically. Additionally, this pragmatic interpretation of knowledge will impact the discussion of historical “facts” that HSS teachers deem integral to their aims and that test developers chose for history tests.

Democratic pragmatist theory assumes that social problems should be dealt with through a process of democratic deliberation. This assumption leads to the assertion that what should be taught (and, it follows, assessed) in the history and social studies classroom is a matter of public concern and should be determined democratically. This colors how one evaluates the relationship between the aims of HSS teachers and state standards and assessments in HSS: It is assumed that there must be a relationship between the two if the test is considered to be democratic. Teachers, along with any other member of a democratic society, have the right to influence the standards and assessments in HSS. In this sense, at a much more abstract level, democratic pragmatism will be used as a framework through which a society can democratize education policy analysis and education policy decisions (Smiley, 1999, 630). Having established a theoretical frame for the study, I now turn to the review of the literature.

Related Literature on Curriculum Orientations in HSS

Critics of HSS education claim that teachers’ orientations are to blame for dismal outcomes – particularly teachers who design content based on students’ interests and pedagogy focused around student group work. These arguments have prompted a “back to basics” movement in HSS education. Many have advocated for increased teacher-directed activity in the social studies classroom, rote memorization of key historical concepts and a concentration on chronological (rather than thematic) units of study.

Advocates of this type of curriculum in the social studies also argue for standardized testing in history and the social studies which, it is assumed, will bolster the rigor and increase student success in the field. Their critiques beg the question, “What is the relationship between the aims of teachers of high school history and social studies (HSS) and state standards and assessments related to HSS?” In an effort to answer this question, I reviewed the literature in two separate fields. To get at how the aims of teachers of HSS have been historically conceptualized and researched in the literature, I reviewed the literature on history and social studies teachers’ orientations in the field. I next reviewed the literature on history and social studies standards and assessments.

The first literature search focused on the aims and orientations of HSS teachers. The search included conceptual and empirical pieces that broadly discussed teachers’ belief systems, but was more specifically a search for pieces that addressed history and social studies teachers’ beliefs. The search included work published since Barr, Barth and Shermis’ landmark 1977 study on teaching orientations in HSS. To that end, specific delimiters were used to cull the research body. Articles from journals that are not peer-reviewed, that were international in scope or dealt solely with elementary classrooms were also excluded. This search returned 58 conceptual and empirical studies. Unless specifically noted, the first section, “Gatekeeping and Orientation in the Field,” summarizes reviews of the literature and empirical studies about how teachers define the purpose of HSS generally and their role as “gatekeepers” in the field. The second section, “The Three Traditions Thesis as a Framework” addresses conceptual studies or

conceptual parts of empirical studies. The following two sections address findings from empirical studies. The final fifth section is my concluding remarks about these 58 studies.

Conceptual Studies: Gatekeeping and Orientations in the Field

An extensive review of the literature on HSS teachers reveals widespread acceptance of the view that the proper aim of social studies is citizenship education (Barth, 1997; Grant & Vansledright, 1996; Lobes, 1998; Longstreet, 1997-98). Four existing reviews of the literature in HSS aptly capture this enduring theme (Marker & Mehlinger, 1992; Mitzel, 1982; Thornton, 1994, 2008). Thornton wrote that “most social studies leaders and policymakers justify the subject on the grounds of citizenship” (Thornton, 1994, 224). Marker & Mehlinger stated that there was “agreement about preparing youth to possess knowledge values and skills needed for active participation.” The agreement about the citizenship function of HSS education is broadly shared by practitioners in the field. This, however, is where the consensus ends.

While most HSS practitioners share the belief that they are preparing future citizens, the definition of a “good citizen” is vague (Longstreet, 1997-98; Marker & Mehlinger, 1992). Stodolsky & Grossman (1995) corroborated this notion and contended that of the five academic subjects in the K-12 curriculum, social studies is among the most poorly defined. The content is enormous in scope drawing from a number of professional fields (e.g. anthropology, history, political science, psychology, sociology). “Perhaps,” they noted, “it is not surprising that diverse political and moral positions bear on the task of defining social studies curricula” (Stodolsky & Grossman, 1995, 231).

In Thornton's (2008) review of the social studies curriculum, he noted that aim of social studies education "can be realized through a variety of means" (p. 15). For this reason, K-12 history and social studies teachers act as, what Thornton called, "curricular-instructional gatekeepers" (Thornton, 1991). That is to say that a teacher filters the curriculum that she presents in her class based on previous (usually implicit) notions of the purpose of social studies.

Wilson and Wineburg's (1988) empirical study revealed that teachers filter the social studies curriculum based largely upon values, norms and conventions found in their previous professional discipline. While all four of the teachers in their study – in one way or another – professed to teach history to create empowered and informed citizens, each went about the task in very individual manners. In the end, Wilson and Wineburg argued, "for our teachers, their 'knowledge' of the subject matter was as much a project of their beliefs as it was an accumulation of facts and interpretations" (p. 537).

In a year-long field study of HSS teachers, Cornbleth (1998) observed teachers in her study portraying varying images of "America." The most consistent image of America the novice teachers portrayed was that of a country that is "imperfect but best" (P. 622). The author argued that the notion put forth by many conservative critics that the nation's school teachers should convey a singular, coherent and patriotic image of America is unlikely to come about. Rather, Cornbleth argued, "many teachers neither accept nor convey images of an unsullied, progressive America" (P. 643).

Finally, in a qualitative study of four practicing teachers, Ooka Pang and Gibson (2001) found that the beliefs, experiences and values that four African American teachers

brought to their classroom significantly affected what was taught in their classroom. The authors found that key themes emerged among and across the four participants' teaching. These themes were Racism and Civil Rights, Responsibility of Citizenship, Social Justice and Slavery. The authors concluded that the four African American teachers centered curriculum on "the experiences of African Americans who are only marginally included in social studies resources," and that they were uniquely able to "engage their students in dialogue about slavery and civil rights" (p. 266).

In each of these studies, researchers found evidence to corroborate Thornton's premise that teachers act as curricular-instructional gatekeepers (Thornton, 1991, 2001). Thornton further notes that a practitioner's consistent ascription to a particular curriculum reflects that teacher's "curricular ideology" (Thornton, 2008, 15) or what I refer to here as "curricular orientation." While one or two curricular orientations may be in ascendancy in any given time period, observers must recognize these multiple orientations in the field when assessing what students learned in these diverse classrooms.

There is significant body of literature in which scholars attempt to categorize various HSS curricular orientations for the purpose of defining a field that has been notoriously vague. I use a framework for discussing these orientations provided by Barr, Barth and Shermis (1977). Next, I review empirical studies that provide a snapshot of how HSS teachers, nationwide, represent each of the curricular orientations and discuss some methodological concerns raised by the empirical studies. In the concluding thoughts I discuss implications for the professional field of social studies.

The “Three Traditions” Thesis as a Framework for Understanding Orientations

In 1977, Barr, Barth and Shermis authored a piece for the National Council of the Social Studies (NCSS) titled, “Defining the Social Studies.” They demonstrated that there is, contrary to most opinions, some consistency between HSS teachers. Indeed, they found that most teacher practitioners ascribed to one of “three separate, conceptually different curricular traditions” (Barr, Barth, & Shermis, 1977, p. 58). They labeled these traditions social studies taught as cultural transmission (CT), social studies taught as social science (SS), and social studies taught as reflective inquiry (RI). Hereafter, I refer to this as the “three traditions” thesis as it has come to be known in the field.

Democratic pragmatists contend that there is room for both the conservative and the liberal in education. Conservative education promotes social cohesion and ease of communication between citizens. Meanwhile, liberal educators work to bring students beyond the status quo. Viewed in this light, the three traditions thesis incorporates both conservative and liberal aims of education with the “reflective inquiry” tradition occupying the left, or liberal-end of the field and the “CT” and “SS” traditions occupying the right, or conservative-end of the field. Rather than assuming, as democratic pragmatists do, that good education should incorporate both the conservative and the liberal leanings of educationⁱⁱⁱ, “Barr, Barth and Shermis devised a framework that assumed that teachers adopt one or the other tradition – that is, teachers will view social studies education as having a conservative purpose (social studies as cultural transmission or as social sciences) or they will view social studies as having a liberal

purpose (reflective inquiry). This distinction will be made clear once the traditions are better defined later in this section.

The authors defined the traditions as a set of beliefs *and* actions – in other words, a teacher’s beliefs about the purpose of history and social studies are linked to the teacher’s curricular and pedagogical choices. Other HSS scholars have since noted that curricular and pedagogical choices are a natural extension of one’s beliefs about the purpose of their field. Marker and Melinger insisted that “...the social studies curriculum depends on one’s perspective” (Marker & Mehlinger, 1992, 833). Thornton also argued that “there is no isomorphic relationship between a conception of curriculum... and a particular pedagogical approach; one is often associated with the other in practice (Thornton, 1994, 224).

Despite the existence of dozens of conceptual and empirical studies that propose modifications to the typology, the new categories resemble very much, or are simply divisions of the existing three traditions framework. In other words, the three traditions thesis has endured (Thornton, 1994, 224). For that reason, I use the three traditions framework to review the literature. Unless noted, the following section is a review of conceptual literature and conceptual parts of empirical studies. In the instances where conceptual parts of empirical studies are included, the research designs of those studies are described in greater detail in the following section of empirical studies. Here, I include the empirical studies for the purpose of surveying the manner in which the labels are defined by scholars in the field. I address outliers at the end of the discussion. So, what are these traditions and how have they since been modified?

Social Studies as Citizenship Transmission (CT)

By far, the most easily defined and readily observed tradition of HSS teaching includes those who view HSS as citizenship transmission (CT). All of the studies published post-Barr includes a category such as this – if not directly referred to as “citizenship transmission” authors label this orientation similarly. For example, Martorella (2001) labeled this practice as “social studies taught as transmission of cultural heritage.” Morrissett & Haas (Ross, 1997) referred to it as “conservative cultural continuity” and Seixas (2000) as the “collective memory” approach. Bennett & Spalding (1992) referred to these practitioners as “inculcators,” Anderson *et al* (1997) as both “legalists and assimilators,” Abowitz & Harnish (2006) called these practitioners “civic Republicans” and Goodman & Adler (1985) as “citizenship indoctrinators.”

As indicated in the name, practitioners in the CT tradition view the primary purpose of history and social studies as inculcating pupils with traditional knowledge. Linking this to the theoretical frame, many would refer to this tradition as culturally conservative – that is that practitioners in this tradition wish to preserve and promote the status quo. In an earlier work, Morrissett (1977) referred to teachers who practice in the citizenship transmission tradition as “historians.” He noted that with these practitioners “History [is] the major and/or integrating focus of study” (p. 206). Brubaker *et al.* (1977) supported this notion and labeled practitioners with a CT orientation as teachers who view “social studies as knowledge of the past as a guide to good citizenship.” Vinson (1998) observed that teachers in this orientation hope their pupils will “acqui[re] ‘American or ‘democratic’ values vis a vis the teaching of discrete, factual pieces drawn

primarily from the canon (p. 56).” Teachers working from a CT orientation, Vinson argued, expose their pupils to canonical works, American heroes and key events in America’s past in the hope that it will serve to develop students’ character and instill in them a sense of patriotism and collective identity.

Believing that a good citizen is one who is obedient, law-abiding and participatory, CT practitioners lean toward teacher-directed pedagogy. Anderson *et al* (1997) wrote that in order to inculcate a sense of patriotism, “assimilationist” practitioners tend toward transmission pedagogies. Goodman and Adler (1985) clarified what is meant by transmission-style classrooms by describing them as classrooms in which students passively memorize American symbols and heroes and the recite information such as the pledge of allegiance. Cornbleth (1982) described the central goal for an “illusory” teacher as teaching students to behave properly and to passively acquire information. Brubaker (1977) added that these students often read aloud and are drilled with information. In the CT practitioner’s classroom, the teacher is the authority and students recite and memorize. Janzen (1995) noted that there is little room in the CT practitioner’s class for student-directed learning. Rather, the curriculum consists of pre-defined viewpoints – collective adherence to these American ideas supposedly leads to a cohesive public.

Pupils who are taught in the CT tradition are trained to become participatory citizens. However, in the CT tradition, civic participation is generally limited to voting. Westhemier and Kahne (2004) described a citizen who has been taught in this tradition as “personally responsible” – that is one who has character and integrity, self-discipline and

responsibility. Other studies (Abowitz & Harnish, 2006; Cornbleth, 1982) argue that CT practitioners expect their pupils to engage in democracy through voting and other mild civic activities such as jury duty – that is the limit, however, to the democratic activity that CT practitioners advocate.

Social Studies as Social Science (SS)

Barr, Barth and Shermis defined the second tradition as “social studies taught as social science (SS).” As with the CT tradition, the SS tradition falls on the right, or conservative end of the field. This tradition is recurrent in the literature. Vinson referred to this as the “social science approach” (Vinson, 1998), but it has been labeled by others as “discovery” (Janzen, 1995), “social studies as structure of the disciplines” (Brubaker, et al., 1977), the “disciplinary approach” (Seixas, 2000), “intellectual aspects of history and the social sciences” (Morrissett and Haas in Ross, 1997) and “subject-centered approaches” (Stanley and Nelson in Ross, 1997). Evans referred in an empirical piece to teachers who adopt a similar perspective as “scientific historians” (Evans, 1990). In a later conceptual piece, he returns to the original three traditions label and called this the “social studies as a social science” camp (Evans, 2006).

As indicated by the labels used to describe this teaching tradition, teachers who work within the SS tradition view the purpose of their discipline as training students to become novice social scientists. Pupils are taught to incorporate techniques used by social scientists in the field (anthropology, archeology, economics, geography, history, political science, psychology, sociology etc.) to consider predetermined social issues and problems. By viewing social issues “through the lens of the professional field” (Janzen,

1995), pupils are introduced to the scientific method and are introduced to social sciences as a career path.

Many scholars of more contemporary pieces muddy the distinction made by Barr, Barth and Shermis between the social sciences (SS) tradition and the reflective inquiry (RI) tradition. However, the distinction between SS and RI is clear. Barr, Barth and Shermis (Barr, et al., 1977) described the SS stance as conservative in structure. They note that students are encouraged to perceive the world “as it actually is” (p. 62) and that guidebooks are used by SS teachers to “transmit the knowledge-gathering ways of the social science disciplines” (p. 63).

A number of other scholars acknowledge the conservative tendency of the SS tradition. In his response to the three traditions thesis, Engle (Barr, et al., 1977) wrote that the SS tradition is consummate to “social science positivism” and that practitioners in this tradition work to preserve the status quo. Janzen (1995) wrote that in the “discovery” tradition, the curriculum is teacher controlled and that the students are encouraged to move toward preselected answers to problems posed by their instructors (p. 136). SS teachers also wish to introduce students to career paths in the social sciences. This has lead some scholars to describe these teachers as having “technical” (Cornbleth, 1982) or “efficiency” orientations (Evans, 2006).

While these scholars’ definitions of the SS tradition are aligned with the earlier Barr, Barth and Shermis work, others connect the SS tradition to the methods of inquiry associated with professional social science fields. In other words, instead of students arriving at predetermined answers, students are encouraged to arrive at their own

conclusions and formulate their own evaluations based on the data much like a social scientist would. In this sense, rather than being a conservative orientation as Barr, Barth and Shermis described the SS tradition, these scholars root the tradition in the more liberal inquiry approach where knowledge is constructed by students. For example, in his empirical study of three history teachers, Evans (2006) described “Rusty” as a “scientific historian.” While the name he ascribes to Rusty’s practice harkens back to Barr *et al*’s “social science” tradition, Rusty encouraged his students to use historical inquiry and their own insight into historical problems to make judgments based on the evidence. This instructor’s orientation speaks more to the reflective inquiry tradition – a tradition in which, as Engle (1977) described it, “inquiry theory, social criticism and policy studies are dynamic theories, they assume change...reform [is] the natural proclivity of humankind (p. 104).” While no authors note the inconsistency in the literature, it appears that those who conflate the SS tradition with more critical, inquiry-oriented traditions may be doing so against Barr, Barth and Shermis’ original intentions.

Social Studies as Reflective Inquiry

Barr, Barth and Shermis’ labeled the third tradition of the social studies the “reflective inquiry” tradition. This tradition falls toward the left or liberal end, of the field of teaching orientations in that, rather than transmitting information, teachers assume that students will conduct their own investigations and arrive at their own conclusions and evaluations based on the evidence. The RI tradition has consistently appeared in the literature on social studies teaching orientations. Over the years, however, the manner in which scholars identify the RI tradition has expanded and splintered. Thus, the tradition

includes an array of labels and descriptions. Scholars appear to describe three different types of inquiry-based positions. Here, I label these different positions “reflective inquiry,” “reflective inquiry for personal ends,” and “reflective inquiry for social ends.” Each label is described in depth below.

Reflective Inquiry (RI)

Contemporary scholars use the label, reflective inquiry (RI), in much the same way as Barr, Barth and Shermis described it three decades ago. Practitioners in this tradition view the purpose of their profession as training students in the inquiry skills needed to become effective citizens. Rooted in the education philosophy of John Dewey, the RI tradition requires that students learn traditional knowledge. Students are then expected to build upon that knowledge to critically assess their world. This teaching orientation is closely aligned with the theory of democratic pragmatism explained in earlier sections of this work in that teachers in this orientation contend that education in a democratic society must liberate the minds of individuals (Rorty, 1999). To do so, individuals must be exposed to multiple experiences and interpretations of the “good life” (Gutman, 1982; Rorty, 1999). Brubaker et al. (1977) explained in this lengthy quote:

Industrial and technological advances in the United States have provided American youth with more and more cultural alternatives. There is a diminution of culturally fixed mores and values – with traditional values, founded in Puritan morality, the work-success ethic, and individualism and achievement being supplanted by emergent values...Given these cultural realities, social studies educators must be responsible for creating conditions whereby students can

inquire into beliefs, values and social policies as well as assess the consequences and implications of possible alternatives (p. 203).

Like the SS tradition, teachers in the RI tradition introduce their students to inquiry techniques drawn from the professional methodologies of the social sciences (anthropology, archeology, economics, geography, history, political science, psychology, sociology). Though they appear similar, one key distinction between the RI tradition and the SS tradition is notable. Rather than expecting pupils to arrive at predetermined conclusions – as is the case in the SS tradition – teachers in the RI tradition assume that students are capable of arriving at their own conclusions. Thornton described this distinction as teaching that is transformative rather than mimetic (Thornton, 1994). The transformative, subjective, even relativistic, stance of Reflective Inquirers runs contrary to the mimetic, fixed, transmission stance of CT and SS practitioners.

Teachers in the RI tradition often treat controversial issues or problems in contemporary society as “starting points” for their curriculum (Janzen, 1995). These problems may be issues that students chose to research (Brubaker, et al., 1977; Evans, 1990; Janzen, 1995; Vansledright & Grant, 1994) but are sometimes problems chosen by the teacher. In the RI tradition, students use research methods to articulate the problem, create hypotheses, devise a method of investigation, collect data, analyze their results and make conclusions based on the data (Brubaker, et al., 1977; Vinson, 1998).

Several social studies scholars discuss the RI tradition in their own work but describe the tradition using different labels. For example, Janzen described the “inquiry” stance as one in which pupils are trained how to think about problems but Morrissett and

Haas (1982) called this “reflective thinking.” Seixas (2000) called this the “postmodern approach” and Martorella (2001) called it “social studies taught as reflective inquiry.”

Finally, Evans (1988, 1990) described these teachers as “relativist reformers.”

Drawing from Cornbleth’s typology of social studies teaching orientations, VanSledright and Grant (1994) used both observation and interview to describe the curriculum and pedagogy of a “constructivist” teacher. Quoting Cornbleth, VanSledright and Grant (1994) wrote that the constructivist teacher encourages students “to pursue their own interests, engage in a variety of activities, and examine a broad range of political content and activities... [Within her classroom,] knowledge is tentative, there are multiple ways of learning and knowing and different perspectives ought to be considered” (p. 330).

In their empirical study, Anderson *et al* (1997) found two groups that were similar to the RI tradition. Those who were considered “critical thinking” practitioners

...reject citizenship education as ... a mechanism for endorsing traditional unreflective understandings of patriotism. Rather, they seem to see citizenship education as a potential way to encourage students to question ‘school and society’ and to raise significant questions about the status quo (p. 348).

Meanwhile, Anderson *et al* (1997) also described “cultural pluralist” educators as those who expose students to various ideologies. These teachers support teaching “political tolerance and open-mindedness and students’ developing an understanding of different cultures [in the United States] and abroad” (p. 350). Both Anderson *et al*’s “critical

thinking” orientation and their “cultural pluralist” orientation resemble aspects of teachers who teach from the RI tradition.

While dozens of studies exist that describe teachers in the RI tradition, many of them do not touch on the pedagogical approaches of these teachers other than to note that they are student-centered rather than teacher-directed, and draw from the inquiry methods of the professional social science disciplines. Some studies offer more depth. For example, Bennett and Spalding (1992) claimed that “facilitators of thinking and lifelong learning” use games, simulations, small group work, and inquiry methods to guide their classroom methods. Abowitz and Harnish (2006) wrote that teachers in the “political liberalism” camp use deliberative methods, discussion and disagreement, and communication and intellectual skills to teach pupils their rights as individuals and to adopt a critical attitude toward authority.

Many practitioners in the RI tradition hope to produce what Westheimer and Kahne (2004) called a “participatory citizen” (i.e., citizens who have the skills to reflect on problems, make decisions and form strategies to remedy those problems). Participatory citizens eventually become leaders in civic organizations and government agencies. Whether these practitioners are referred to as “relativist reformers,” “reflective inquirers” or “political liberals,” teachers in this orientation train their students – and subsequently future citizens – to thoughtfully inquire into, reflect upon and critically analyze social problems.

The RI tradition then is rooted in the practitioner’s desire for students to learn the skills of inquiry. The common goals, curriculum and pedagogy of Reflective Inquirers

seem to end there. And this is where Barr, Barth and Shermis' three traditions thesis falls short. Brubaker (1977) claims that two of the critiques of those who view social studies as "reflective inquiry" are that students should be expected to "go beyond thinking to acting on behalf of their beliefs" and that "not enough attention is given to the emotional side of learning" (p. 204). Indeed, these critiques of the RI tradition are noteworthy in that many practitioners claim that building inquiry skills in youth is not an end in itself but rather it is a means to a much larger end. For these practitioners, the skills of inquiry should be used for one of two purposes: students can use "inquiry for personal ends," or students can use "inquiry for social ends."

Reflective Inquiry for Personal Ends (RIPE)

One way that social studies educators define the purpose of their field is that, as educators, they are responsible for helping children better understanding themselves. Here, reflective inquiry for personal ends (RIPE) is used as a catchall phrase that encompasses several other labels including Janzen's (1995) "life adjustment" stance, Brubaker et al.'s (1977) "student-centered tradition," Goodman & Adler's (1985) "human relations" orientation, Bennett and Spalding's (1992) "scholar psychologist" and "nurturer" orientations, Vinson (1998) and Martorella's (2001) "personal development" approaches, and Morrisett's (1977) "experience" approach.

Similar to Barr, Barth and Shermis' RI tradition, those who view the purpose of social studies as teaching "reflective inquiry for personal ends" (hereafter: RIPE) use inquiry as a stance to investigate important, contemporary social problems and allow students an opportunity to reflect upon how those problems are interpreted by students.

RIPE practitioners differ from the RI tradition, however, in several key ways. First, RIPE practitioners see the inquiry stance as a means to “teach children about themselves” (Goodman & Adler, 1985) and to “help the child comprehend his life” (Janzen, 1995). Sometimes described as humanists, or teachers who focus on values clarification, or character education, RIPE practitioners stress the developmental growth of their students and their role in helping students achieve “a positive self-concept and a strong sense of personal efficacy” (Martorella, 2001; Vinson, 1998).

In terms of content, Bennett & Spalding (1992) noted that “nurturing” teachers tend to “deemphasize academic knowledge” (p. 270) and focus instead on building relationships with their students, teacher-student interaction and caring relationships. Goodman and Adler (1985) echoed this notion and claimed that teachers who stress “human relations” focus on cooperation and thematic units based on communication and problem solving rather than academic or “substantive” historical content knowledge (p. 8).

Reflective Inquiry for Social Ends (RISE)

Another shortcoming of the Barr, Barth and Shermis’ (1977) description of the RI tradition is that it fails to include educators who explicitly teach their history and social studies students to become change agents. Subsequent literature has defined more accurately teachers who adopt this mission-type stance to HSS instruction. The literature has described transformative teachers as those who teach social studies for “social action” (Goodman & Adler, 1985; Janzen, 1995), “socio-political involvement” (Brubaker, et al., 1977) or simply “involvement” (Morrissett, 1977), “social transformation” (Stanley,

2005), and “social criticism” (Vinson, 1998). These practitioners are also referred to as “social meliorists,” “social reconstructivists,” “critical pedagogues” (Evans, 2006) and “empowerers” (Bennett & Spalding, 1992).

Again, reflective inquiry for social ends (hereafter: RISE) practitioners maintain similar beliefs as RI instructors in that inquiry techniques play a central role in the development of their curriculum and instruction. Yet, in addition to this inquiry stance, RISE practitioners focus on problems that stimulate and/or increase students’ social awareness (Goodman & Adler, 1985). By focusing on social problems or incidents of historical injustice, RISE practitioners provide students with the opportunity to examine and critique current social practices and policies (Brubaker et al., 1977; Evans, 2006; Martorella, 2001; Vinson, 1997). These practitioners expect their students to become experts in problem solving (Martorella, 2001; Vinson, 1997). Furthermore, graduates of their classes are expected to be, what Westheimer and Kahne (2004) refer to as “justice oriented.” That is to say that RISE practitioners expect their students to use problem solving capabilities to actively transform society (Bennett & Spalding, 1992; Brubaker et al., 1977; Janzen, 1995; Martorella, 2001; Morrissett, 1977; Vinson, 1997).

Outliers to the Three Traditions Thesis

As stated previously, despite the fact that Barr, Barth, Shermis’ three traditions framework is over three decades old, the framework continues to be an appropriate heuristic for understanding HSS teachers orientations in the field. However, subsequent studies have found a handful of what can only be described as “outlier” teaching orientations or approaches in the classroom.

Goodman & Adler's (1985) empirical study provides two examples of such outliers. Using data gathered throughout one year of observations and interviews, the authors found that the 16 teachers in their study viewed social studies in six different ways. Four of these categories (human relations, citizenship indoctrination, school knowledge and social action) are readily identifiable as categories that bear resemblance to one of the three traditions, and have been discussed previously. Two remaining categories, "social studies as a non subject" and "social studies as integrative core of the elementary curriculum" do not fit neatly into the three traditions thesis. This is due largely to the fact that Goodman & Adler's study focused on the perspectives of elementary teachers – many of whom have some level of curricular autonomy as to how HSS subjects are included in the curriculum. For that reason, these teachers are labeled not only based upon how they believe HSS should be addressed in the curriculum, but *how often*.

Bennett and Spalding's (1992) empirical study drew from data gathered over four-year, qualitative research project that found seven different teaching perspectives among 68 individual teachers of all subjects. The authors found that HSS was the only academic subject in the study that had a representative in each of the seven perspectives. That is to say, the social studies teachers were the most diverse in terms of how they viewed the purpose of their field. More than half of the authors' teaching orientations (the inculcator, the facilitator of thinking and lifelong learning, the empowerer, and the nurturer) align with the three traditions. However, three of the authors' teaching

perspectives (the friendly scholar, the friendly pedagogue and the scholar psychologist) do not fit neatly into the traditions framework.

As with the Goodman and Adler study, this schema differs from the traditions framework in that where the traditions framework describes HSS teachers' beliefs about the purposes, curriculum and pedagogy of HSS, Bennett and Spalding's schema describes additional phenomena. Both the "friendly scholar and friendly pedagogue" teaching perspectives describe how the practitioner views their relationships with students in their classroom. Meanwhile, the "scholar psychologist" category is something of a catchall category. Indeed, the authors describe scholar psychologists as the largest and least clearly defined group in the study – its subscribers tended to take on the characteristics of many previously described perspectives.

A final set of outliers is found throughout the work of Evans (1988, 1989, 1990, 2006). In two separate exploratory investigations of classroom teaching, Evans described teachers' conceptions using five categories. Three of these conceptions (the storyteller, the scientific historian and the relativist reformer) are moderately aligned with the existing traditions framework, though some clear disconnects between the two surfaced. This was noted in previous sections. Two of Evans' conceptions (the cosmic philosopher and the eclectic) stand apart entirely from the traditions framework. The "eclectic" conception does bear some resemblance, by and large, to Bennett and Spalding's "scholar psychologist" category in that, as the name suggests, an "eclectic" draws from multiple conceptions of history to inform his or her practice. However, Evans' "cosmic philosopher" conception is a true outlier in that it resembles no other orientation found in

the review of the literature. The cosmic philosopher, as Evans' described it, focuses on the connections between history and philosophy. A cosmic philosopher highlights the repetitive nature of history and connects history to larger concepts and themes outside of the discipline. While Evans' did cite the Barr study among his references, he made no connections between his findings and extant theory on teachers' orientations.

In a later conceptual piece (2006), Evans described "five major competing camps" of social studies curriculum orientations. None of the five conceptions Evans' described in previous empirical studies appears in this new typology. Evans labeled the first two camps as "traditional historians" and advocates of "social studies as social science." These two categories are clearly in line with Barr et al's CT and SS traditions. Yet, mysteriously Evans described the final three camps as "social efficiency educators" "social meliorists," and "social reconstructionists or critical pedagogues." While those he described as social efficiency educators can easily be viewed as practitioners in the CT tradition, social meliorists, social reconstructionists and critical pedagogues are virtually indistinguishable in the traditions framework. The argument can be made, as Abowitz & Harnish (2006) did, that those in the RI, or liberal tradition, differ qualitatively from social reconstructionists in that reconstructionists call for a more extreme revision of the status quo than do reflective inquirers. Also, rather than describing current orientations in HSS teaching, Evans described the historical "pendulum swings...of the curriculum landscape" (Evans, 2006, 317). One can only assume that Evans' typology is based largely on earlier work on curricular struggles in the general curriculum (Kliebard, 2004) but are therefore, not directly connected to HSS purposes, content and pedagogy.

Regardless, the disconnect between Evans' work and earlier theoretical work in this field made Evans' work particularly difficult to include in the three traditions framework. His entire body of work seems to exist as an outlier to previous scholarship.

Findings of Quantitative Empirical Studies: Demographic Differences

Barr, Barth and Shermis' three traditions (CT, SS, RI) including the two sub-traditions (RI for Personal Ends and RI for Social Ends) listed above provide a useful heuristic for understanding the types of teaching likely to be found in classrooms throughout the United States. It remains to be seen which of the traditions most accurately defines the majority of practitioners in the field. For example, many scholars argue that the CT tradition, while the least supported tradition in professional scholarship and teacher preparation programs, consistently remains the most frequently observed tradition in practice in classrooms. What can be said about the prevalence of each of these teaching traditions in HSS classrooms throughout the nation?

Since the original Barr, Barth and Shermis' study based upon the "Social Studies Preference Scale" confirmed the existence of the three traditions, four additional empirical studies have since emerged in which scholars employed large-scale survey methods to assess the prevalence of specific teaching orientations in the field (Anderson, et al., 1997; Leming, Ellington, & Schug, 2006; Morrissett, 1977; Vinson, 1998).^{iv}

Moving in chronological order, Morrissett's (1977) Curriculum Information Network study of "Preferred Approaches to the Teaching of Social Studies" surveyed a self-selected sample of 440 social studies practitioners at the senior and junior high level, college teachers, consultants and supervisors. Morrissett's survey found the vast majority

of respondents identifying with the “critical thinking” approach (similar to the RI tradition). Respondents were less likely to choose the SS approach as their preference and were least likely to chose “history” – the approach most resembling the CT tradition.

Anderson et al (1997) used a Q-technique to discern the “principal conceptions” held by social studies teachers in a small sample. They then devised a survey based upon descriptors of these four categories and administered the survey to a national random sample of 800 members of the National Council of the Social Studies (NCSS). Anderson et al. received 361 completed surveys – a response rate of 45%. The authors were able to find a statistically significant difference between their four perspectives. The somewhat more liberal perspectives – critical thinking and cultural pluralism – dominated (p. 352). Not surprisingly, the researchers were able to link particular teaching perspectives with certain demographic characteristics. For example, those who viewed citizenship education from the critical thinking and cultural pluralism perspective were more likely to be Democrats and to teach high school. They also tended to live in certain regions in the country including the Pacific states and New England. Those who adopted the more culturally conservative assimilation perspective tended to be from small cities and towns and were more often Republican and conservative in political orientation.

Vinson’s (1998) empirical study was based on a survey of 500 high school social studies teachers and members of the NCSS. His survey was comprised using Barr, Barth and Shermis’ original Social Studies Preference Scale that looks for indications of the three traditions. Vinson added items to this survey to reflect Martorella’s two additional categories: social criticism and personal development. The participants in this study

identified strongly with the liberal leaning approaches of social studies as reflective inquiry (RI), social studies as informed social criticism (RISE) and social studies as personal development (RIPE) than they did with more conservative approaches.

Finally, Leming et al.'s (2006) national random survey of elementary and middle school social studies teachers drew from data collected via telephone of 1051 second-, fifth-, and eighth-grade teachers from throughout the nation. The researchers went to great lengths to use all social studies teachers across the nation as a population frame rather than drawing their population sample, as the other studies discussed above did, from the NCSS. Leming et al. did not devise categories of "curricular emphasis" based upon the three traditions thesis. Rather, the authors simply noted that "Teachers were given six rationales often cited for including social studies in the school curriculum" (p. 324). The teachers were asked to rate each based upon the degree of emphasis they felt the rationale should receive in the curriculum. Furthermore, this study did not require teachers to align with one category over another. Teachers could rate each category with equal levels of importance.

The study found that 85% of teachers rated "promoting the acceptance of cultural diversity" with an 8, 9, or 10 – meaning extremely heavy emphasis. 84% rated "the Constitution and U.S. system of government" with extremely heavy emphasis. Meanwhile only 70% believed teaching about "injustice in the American system, with particular attention paid to race, gender and class injustice" should be extremely heavily emphasized in their curriculum. These findings suggest that this sample of elementary and middle school teachers were likely to support some aspects of Barr, Barth and

Shermis' conservative CT approach, and some aspects of the more liberal RI approach, simultaneously.

Several methodological limitations present themselves with the aforementioned quantitative studies. First, two of the four studies mentioned used the NCSS membership list as a population frame from which the researchers drew their random sample. Leming (2003; 2006) noted that while this population frame is easily accessible to and convenient for researchers, members of the NCSS represent only a fraction of American social studies teachers – a population that is not an accurate representation of social studies teachers nationwide. Members of NCSS tend to be more liberal in political orientation than other social studies teachers thus leading to results that favor left-leaning orientations. It should also be noted that members of the NCSS tend to have been professionally trained in teacher preparation programs and have attained higher levels of education in general (Wade, 1993). Leming (1992) notes that these teachers are likely to be taught by professors who support constructivist, inquiry-based and justice oriented teaching. Members of the NCSS are also more likely to endorse and practice research-supported “best practices” of social studies teaching (Myers, Adler, & Brandhorst, 2002). This skewed the results of surveys toward more inquiry-based teaching approaches.

Additionally, some post hoc validation studies (Andres, 1982; Vinson, 1998; White, 1982) found significant overlap between and among the three traditions – that is to say, the practitioners who aligned with one tradition may also have displayed characteristics of another orientation. White's (1982) validation study of the Social Studies Preference Scale (SSP Scale) found that a panel of nine social studies experts,

including four professors and five doctoral students, was able to sort the 45 item SSP Scale into the three traditions with appropriate goodness of fit. Thus the panel was able to establish reliability and content validity for the Traditions. However, additional tests to determine the construct validity of the SSP Scale suggested that there were two, rather than three, traditions. Specifically, White used factor analysis to generate three factors with varimax rotation. He observed that the dimensions represented by the factors validly represented the three traditions but did not represent the purpose, method, and content dimensions of the SSP Scale. White also observed that while the items that the scale's designers intended to measure the CT tradition did so adequately, a third of the items the scale's designers intended to measure the SS and RI tradition failed to load properly with those factors. This, the authors suggested, demonstrated that the scale had only "tapped two dimensions – a CT dimension and a second dimension characterized by elements of both the RI and SS traditions" (p. 12).

Furthermore, White's (1982) study found that the SSP Scale's distinction between "purpose, method and content" was weak. This weak distinction between purpose, method, and content was demonstrated with reliability tests by the expert panel as well as with additional varimax analysis that forced the nine factors onto the purpose/method/content dimensions. White wrote that this analysis "yielded a pattern of item loadings unrelated to the pattern intended by the scale's designers" (p. 12).

Finally, White conducted a correlation test with another construct and matched participants' outcomes on the Barth-Shermis SSP Scale to the Rokeach dogmatism scale to measure construct validity. White theorized that participants identified by the SSP

Scale as “CT” instructors would also test as “highly dogmatic” on the scale of Dogmatism while those participants identified by the SSP Scale as “RI” instructors would test as “less dogmatic.” While the construct match between the SSP Scale and the scale of Dogmatism seems, at least to this researcher, highly questionable, White found that while CT instructors did prove to be more dogmatic, little distinction was found between the RI and SS traditions. White concluded that the SSP Scale seems to measure *two* reliable traditions – those being the dogmatic - CT tradition and the second being a mix of less-dogmatic SS/RI traditions.

In a more recent verification study, Vinson (1998) found moderate levels of both reliability and validity with a modified version of the SSP Scale. Vinson conducted a series of within-category correlations for each instructional approach and found a relatively modest degree of validity. Additionally, he used a simultaneous cross-tabulation procedure to find a dependent relationship between the variables. A discriminant analysis indicated that “selected instructional approach correctly predicted calculated instructional approach for 73.33% of the pilot sample” (p. 67). Vinson also used test-retest procedures across two pilot administrations and a cross-tabulation procedure for selected and calculated instructional approach categories to find moderate reliability of the instrument. Many of Vinson’s findings are discussed in greater detail in chapter three.

Findings from Qualitative Empirical Studies: Teachers as Gatekeepers

Leaving aside instrument validity and reliability studies, these survey study results highlight an interesting divide in the literature on social studies teaching – a divide

that is consistent with existing research on teaching. While results of quantitative survey studies suggested that most HSS teachers aligned with liberal orientations of teaching *in theory*, qualitatively-based empirical studies suggested that teachers did not align with these orientations *in practice*. These studies engaged small samples of social studies teachers and were, therefore, unable to generalize to the nation-wide population of HSS teachers. Yet, the studies described classroom practices of HSS teachers, and many of these studies found a philosophy-to-practice divide.

For example, VanSledright and Grant (1994) mined data gathered through previous observations and interviews of three teachers. Viewing the data through the lens of Cornbleth's (1982) models of citizenship education (the "illusory, the technical and the constructive"), the authors conducted classroom observations and interviews to analyze the practices of teachers who claimed to align with particular teaching orientations. And, while several aspects of one of their participant's teaching "assumed the form of the constructive [RI] approach" (p. 331), the authors note, even in this teacher's classroom, technical and conservative dimensions also emerged. The authors concluded that a teacher's tendency to turn their back on constructive or inquiry-based beliefs and develop more technical and transmission practices is one of the many ways that "persistent classroom teaching dilemmas" impact a teacher's practice.

Evans (1990) observed a similar phenomenon in his exploratory investigation of five classrooms, each representing one of the five typologies discussed earlier (the storyteller, scientific historian, relativist/reformer, cosmic philosopher, and eclectic). Evans noted that while two of the five teachers described what they did in their classroom

accurately, three of the five teachers showed contradictions among what they said they did and what he observed. Evans wrote that, “two of the five teachers spent a good deal of their time and energy managing student behavior problems.” “Thus,” he writes, “my initial impression was one of disappointment” (p. 105). As was the case with the VanSledright and Grant study, Evans observations of a teacher who claimed to be closest to the “reflective inquirer for social ends” stance was furthest from that in his classroom practice. Evans participant, Jeff, was bogged down in classroom management issues. His students rarely listened to, or cared about his political beliefs, and his teaching had little impact on student beliefs. Evans concluded harshly, “One wonders what impact a more gifted teacher might have on students, given a similar typology” (p. 118).

Wilson, Konopak and Readence (1994) mixed methods study used a survey to distinguish secondary pre-service teachers’ conceptions of teaching and learning combined with case study data to see how beliefs informed practice. The authors noted that the case study of David showed a troublesome relationship between the teacher’s beliefs and his practices. Most notably, David stated that he would use alternative approaches to instruction including simulations, role playing and hands-on activities with primary sources. In practice, however, David rarely utilized these teaching techniques. He explained that “[t]hose activities take too much time” (p. 375). Additionally, David voiced his belief in using high-level thinking questions in classroom assessments. Yet again, David’s desire for efficiency led him to put these goals aside noting high level questions meant “the sheer volume [of grading] was too much” (p. 375). The longer

David practiced, the more he employed literal-level questions that were quickly and easily graded.

Concluding Thoughts on History and Social Studies Teachers' Orientations in the Field

These findings are hardly surprising. Existing research on teaching in general (Clandinin & Connelly, 1995; Goodman & Adler, 1985; Lortie, 1975; Wideen, Mayer-Smith, & Moon, 1998; Zeichner & Liston, 1987), and teaching history and social studies in particular (Grant, 2001; Johnston, 1990; Leming, 1992; Metzger & Marker, 1992; Pryor, 2006; Ross, 1987; Yon & Passe, 1994) has demonstrated over time that teachers' beliefs are resilient and are impacted only marginally, if at all, by teacher preparation programs and/or policy mandates – especially programs that work in cross purposes to a teacher's belief system. Furthermore, the reality of the classroom, and a teacher's socialization by colleagues and students to the “grammar of schooling” tend to redirect novice teachers' deeply held beliefs (Tabachnick & Zeichner, 1984; Tyack & Cuban, 1997). Strangely, while critics of constructivist, inquiry-based social studies education contend that the emphasis on the reflective inquiry tradition in teacher preparation programs is to blame for pupils “abysmal” performance on history assessments, research tends to suggest that HSS teachers, in practice, teach the subject in the very manner that the critics advocate – that is through traditional, chronological, cultural transmission (Levstik, 2000; Levstik & Tyson, 2008; Marker & Mehlinger, 1992; Thornton, 1994, 2008; Vogler, 2006).

While teachers tend not to be swayed by their professional preparation programs and have remained relatively conservative in practice over the years, scholars in history

and social studies tend to endorse the opposite. Research on HSS teaching and teacher preparation is permeated with position papers as to what HSS should be, and how teachers should be prepared to meet those goals. Dozens of pieces found through this literature review are simply advocacy pieces meant to sway practitioners and teacher preparation educators to one orientation or another^v. Whether an HSS teacher will adopt the stance advocated for in these position pieces largely depends on their original teaching orientation. For example, a teacher who tends toward the CT tradition is probably far less likely to read and adopt social justice, or critical multicultural stances in their practice. So, while both the professional literature and HSS teacher preparation programs seem highly susceptible to trends (Marker & Mehlinger, 1992; Thornton, 1994), based on the research on teaching beliefs, HSS teaching practitioners seem far less susceptible in theory and in practice to those trends.

What likelihood is there, then, of defining the orientations of history social studies teachers? The majority of the literature reviewed in the preceding pages, including the landmark work by Barr, Barth and Shermis, attempts to do just that – to define a field in unifying terms based mostly upon how professionals define their own beliefs about the purpose of HSS. In his comments on Barr, Barth and Shermis' 1977 work, Engle (1977) wrote that the authors

...have failed to demonstrate convincingly that the unity they sought is either feasible or desirable. Rather, they have succeeded, to an impressive degree (and possibly inadvertently), to clarify the lines of battle between the competing

philosophies... and to point up the irreconcilable nature of the issues which separate them (p. 103).

Marker and Melinger (1992) also conceded that there is little hope the HSS curricular dilemma will ever be resolved. It is likely that HSS professional teachers will continue to teach based upon a relatively fixed belief system as to the rightful purpose, method and content of their field.

To say that these different belief systems may never be consolidated into one unified conception of HSS is not to say they should not be researched and clarified. In fact, the findings point to the need to help pre-service, novice and practicing teachers clarify their teaching theories and think about how these orientations impact both their practice and student outcomes. A few scholars (Eisenhart, Shrum, Harding, & Cuthbert, 1988; Zeichner & Liston, 1987) have highlighted the important role that teacher preparation programs can have in helping teachers to reflect upon, and clarify, their beliefs about classroom practice. Still others (Fickel, 2000; Goodman & Adler, 1985) specifically have called on social studies teachers to consider how their orientations impact what is taught and what is learned in the HSS classroom. These teachers must be taught to reflect upon their own gate-keeping powers. In an era when high stakes testing advocates are increasingly turning their attention to history and the social studies, teachers must consider how their particular orientation may help or hinder their pupils' performance on these tests. Likewise, testing advocates and policy actors must consider the diversity of teaching orientations in the social studies as they call for and design high stakes tests in history and social studies. In order for all students to be given an

opportunity to demonstrate what they have learned in their respective HSS classrooms, testing instruments must reflect the diversity of teaching and learning found therein.

Related Literature on Standards and Assessments in HSS

As noted previously, critics in the field of education have called on HSS teachers to refine their practice so that students' performance on outcome assessments will improve. This begs the questions, what performance standards are pupils currently expected to meet? In what manner do teachers and administrators assess whether students have met these standards? In the following section, I review the literature on standards and assessments in HSS. First, I review the conceptual literature on HSS standards and the conceptual literature on HSS assessments. In the next section, I review the empirical studies on standards and assessments. Unlike the review of the conceptual literature, standards and assessments are discussed together in the review of the empirical research, as these topics are often treated together in the studies.

The review focused on work published since 1991 (the publication date of the last *Handbook of Research on Teaching the Social Studies*), and included conceptual and empirical pieces that dealt specifically with standards and assessments as they pertain to the subjects of history and social studies. With the exception of a few landmark studies (Abrams et al, 2003; Cimbricz, 2002; Madaus, 1988; Pedulla et al., 2003; Ravitch & Finn, 1987) articles that appeared before 1991, or that dealt with the effect of standards and testing on the curriculum *writ large* are excluded from this review.

Conceptual literature on standards and assessments in HSS

The literature search returned 22 conceptual articles on standards and assessment – seven of these focused on standards, and fifteen centered on assessments. The vast majority of these articles dealt specifically with the subject of HSS. While scores of conceptual articles exist that discuss the impetus for, and impact of, standardized tests across the school curriculum in general, for the sake of organization, only one article (Madaus, 1988) was included in this review of the literature. This article was included because of its standing in the education research literature.

Conceptual Studies of HSS Standards

Several key themes consistently appeared across the seven conceptual pieces on HSS standards. The most prominent theme was to do with the enormous scope of content included under the umbrella of social studies. For example, Thornton (2001) remarked that standards makers often confuse the distinction between the “content” (i.e. all of the possible tidbits of knowledge deemed valuable in a particular area), and “subject matter” (i.e. specific content chosen by the teacher for study in a classroom). In their effort to delineate specific subject matter they believe should be taught in the classroom, standards makers mistakenly draw up exhaustive lists of disciplinary content.

The literature reveals that, in the field of American history alone, the content that standards makers choose from is “practically infinite” (Thornton, 2001). The delineation of standards in history therefore creates, what Saunders (1996) called, an “overemphasis on mundane trivia.” Foster and Morris (1991) wrote that the resultant National Standards in History therefore were “bloated, unrealistic, burdensome, and broad” and emphasized

coverage of everything “from Plato to NATO.” While most scholars consider American history the flagship discipline of the social studies, additional content from other disciplines is included under the “social studies” umbrella (Nelson, 1998). Standards in social studies do not just include history concepts but also include content from the disciplines of economics, political science (civics or government), and geography. When the disciplinary content from these professional fields is added to the social studies standards fray, Thornton’s (2001) contention that the standards are “practically infinite” becomes almost an understatement.

A second theme in the literature is how the immensity of the social studies standards impacts teachers at the classroom level. Some of the conceptual pieces note that the sheer volume of content overwhelms teachers (Saunders, 1996; Schneider, 1995). Saunders (1996) wrote that each historical era the standards address has become “a virtual blizzard of random learning objectives and achievement examples that will overwhelm even the most energetic and best-prepared teacher” (p. 65). Still others highlight the fact that the cumbersome nature of the standards runs contrary to the ubiquitous “ethic of practicality” (Foster & Morris, 1991) in the teaching profession – that is to say that teachers generally view the standards as impractical for classroom life (Foster & Morris, 1991; Saunders, 1996).

More importantly, scholars argue that the focus on mundane historical trivia encourages teachers to adopt pedagogical approaches (e.g. “breadth over depth,” “drill and kill,” “puppet from the pulpit,” “chalk talk”) that are considered ineffective at best, and counter-educative pedagogical practice at worst. They argue that the social studies

standards “deprofessionalize” teaching, and deskill teachers, by pushing teachers to transmit content rather than transform pupils (Evans & Pang, 1995; Nelson, 1998; Thornton, 2001). Others argue that the focus on historical concepts rather than on historical thinking skills means that while teachers and students refer to the subject as “history,” professional historians would never concede to apply the same name (Foster & Morris, 1991; Nelson, 1998). Indeed, Foster & Morris (1991) described the history standards as “anathema to the essential tenets of the historical discipline” (p. 65).

A final theme that pervades the conceptual literature on HSS standards has to do with the political nature of defining standards. Power struggles over the curriculum are neither new, nor exclusive to history and social studies (Pinar, 1996). Yet, defining what is to be taught about the past, present and future of the United States’ certainly draws a fair share of the public’s attention. Interested parties seem to have wildly varying opinions as to what American students should know. Oftentimes, these divisions are drawn along political party or disciplinary lines such that each standard turns out to be “intractably dogmatic, controversial and problematic” (Foster & Morris, 1991). The battle over the curriculum becomes particularly contentious when it comes to multicultural concepts (Foster and Morris, 1991; Schneider, 1995). A long standing debate among advocates of multicultural education centers on teachers’ tendencies to employ additive, rather than critical narratives when discussing the role of people of color in American history (Banks, 2004; Cornbleth, 2006; Evans & Pang, 1995; Schneider, 1995; Sleeter & Grant, 1987). While standards-makers in some social studies disciplines argue that they have worked to include multicultural concepts (Evans, 2006; Nash,

Crabtree, & Dunn, 2000), proponents of critical multicultural education attest that adding to the curriculum names, biographies, dates and other facts meant to represent diverse voices does nothing to encourage students to critically examine society or to work for social justice (Banks, 2004; Sleeter, 1995, 2000; Sleeter & Delgado Bernal, 2004).

Conceptual Studies of HSS Assessments

Several themes emerged across the 15 conceptual pieces on history and social studies assessments. Chief among the conceptual pieces, as mentioned previously, is Madaus' (1988) groundbreaking article, "The influence of testing on the curriculum."^{vi} In the piece, Madaus outlined seven principles that describe the effects of standardized assessments on the curriculum – principles that, in the twenty years since the publication of this piece, have risen to the level of truism in discussions about standardized testing. The principles in one way or another appear throughout, and perhaps even define, the conceptual literature on standardized assessments. Below, I examine the conceptual literature on assessments in history and social studies using Madaus' principles as a framework with two slight modifications. First, Madaus' principles are discussed in his article in a particular order but I begin with Principle 2, move in order and end with Principle 1. Additionally, rather than discussing Principle Four and Five separately, I treat them as a unified set of principles. These changes were made for organizational purposes. Throughout, I note how (little) the principles have evolved in the literature, and how each applies to HSS specifically.

Principle 2: If important decisions are presumed to be related to test results, then teachers will teach to the test.

In 1988, Madaus claimed this was one of many ways that assessments impacted the school curriculum. Since then, the adage “what gets tested, gets taught” has become an oft-repeated mantra of testing critics. Yet, this assumption, when applied to the discipline of history and the social studies, reveals that the maxim is true in the negative as well – what does not get tested, does not get taught. Indeed, in districts where schools are working to meet adequate yearly progress required by the *No Child Left Behind Act* (hereafter: *NCLB*), the subjects of English and math supersede history and social studies in level of priority. Kurfman (1991) noted that when social studies is not included in school testing programs, there is evidence that the subject receives little time. This is particularly true in elementary classrooms (p. 313). Volger (2003) wrote that when his principal explained that they needed to make more time for English and math because a high percentage of their students had failed the Massachusetts Comprehensive Assessment System (MCAS), he begrudgingly understood her logic. Whether or not it is empirically proven to be the case, surveys show that teachers *believe* that unless social studies is tested, it will not receive adequate attention in the curriculum (Grant, 2006; Horn, 2006; Savage, 2003). For this reason, Savage (2003) wrote, many teachers reluctantly advocate for testing in history and social studies simply so that it is given priority.

Meanwhile, in states that have in place, or plan to introduce, state-wide standardized assessments in history and social studies (not required by the *NCLB*

legislation), a hierarchy of content-area importance develops with HSS among the top competitors (Volger, 2003). While the effect of state-mandated tests on HSS curriculum and pedagogy will be explored further in the section on empirical literature, it may be noted here that many HSS teachers view this greater prominence – and subsequent pressure to help their pupils to perform well on the test – as the lesser of two evils. One scholar refers to this phenomenon – in which HSS teachers are damned if they test, and damned if they don't – as “the social studies squeeze” (Burroughs, Groce, & Webeck, 2005).

Principle 3: A high-stakes test transfers control over the curriculum to the agency which sets or controls the exam.

Testing critics have long complained that using tests as a lever to reform curriculum and pedagogy in the schools is akin to having the tail wag the dog (Kurfman, 1991, 317). They assert that while the public has a right to know how well schools are doing (Savage, 2003, 202), turning to testing agencies transfers authority over the curriculum to people at the upper levels of government who are “far removed from the actual classrooms” (Pahl, 2003b, 197). Test developers, rather than curriculum specialists, are now at liberty to determine what qualifies as a quality social studies program. Allowing test developers to define history and social studies through the tests they develop, Savage (2003) argued, will have significant consequences for the profession.

Primarily, scholars claim that classroom teachers (who are closer to the pupil and are arguably better suited to make assessments as to the child's ability in a subject area)

have been brushed aside by standardized tests. Gaudelli (2002) contended that teacher decision making is a fundamental, and complex, pedagogical process – particularly in a field like social studies, which includes innumerable concepts and insurmountable philosophical divisions. Several other scholars pick up this thread by noting that the diversity of pedagogical and curricular orientations makes it nearly impossible to accurately align standardized assessments with what is actually taught in the classroom (Kurfman, 1991; Pahl, 2003a, 2003b, 2003c; Volger, 2003). Pahl (2003a, 2003b, 2003c) contended that while HSS assessments are meant to measure the acquisition of subject matter taught throughout the academic year, tests are often administered in mid-spring creating a situation where students are being tested on subject matter that will not be taught for another month or two. Citing previous studies, the author noted that 80% of questions on one ninth-grade state-wide social studies exam were not part of the curriculum for, and therefore not valid for the state’s ninth-graders. Pahl (2003a) called on testing agencies to create oversight committees drawn from professional educational and social studies organizations to “guard against unfair and inept tests and to ensure the inclusion of better test items on standardized tests” (p. 215). Gaudelli (2002) concluded that because of gross misalignment, and because the field is so “fundamentally fragmented” (p. 200), the professional community should not treat standardized tests results in HSS as an accurate assessment of pupils’ knowledge acquisition.

Despite the fact, as argued previously, that Maduas’ principle of “transferred control” has risen to the level of truism in the conceptual literature on HSS standardized testing, several scholars have called for more empirical research to determine if this is

indeed the case for social studies. Two scholars (Grant, 2006; Kurfman, 1991) wrote that the notion that tests drive curriculum, or act as a lever of change, has never been validated by empirical studies. Further research is needed to determine both if, and how, tests drive curricular change. If it is indeed found that tests, rather than teachers, dictate the curriculum, Kurfman (1991) called for further research to examine the effect on teachers' professional morale.

Principle 4 & 5: In every setting where a high-stakes test operates, a tradition of past exams develops, which eventually de facto defines the curriculum. Teachers pay particular attention to the form of the question on a high stakes test (e.g. short answer, essay, multiple choice), and adjust their instruction accordingly.

In Madaus' landmark article, the author treated the principle listed above as two separate principles. Here, one principle is treated as the natural extension of the other. A pervasive theme in the conceptual literature on history and social studies assessments holds that tests "define the curriculum" in myriad ways. Scholars argue that tests first define the content that is taught and secondly they define the methods teachers use to assess their pupils. This relationship between content and assessment tools is deeply interrelated, symbiotic and cyclical.

First, scholars argue that standardized HSS tests define the curricular content in HSS classrooms. They contend that teachers take note of topics that are tested often—those topics are then preserved and highlighted in the curriculum (Pahl, 2003a). Rees (2003) argued that local history, social history and multicultural history are often dropped from the curriculum in favor of more conservative, "consensus history," which

unfailingly focuses on the progressive march toward a better society rather than the enduring conflicts that plague contemporary America.

Scholars are quick to point out that when HSS topics are tested, teachers are encouraged to move through content in a fast-paced and shallow manner. Due to the enormous scope of social studies, and the random nature by which a concept is chosen to be an item in an assessment, an inverse relationship has developed between standardized assessments and curricular depth – that is, if a test is present, depth disappears (Rees, 2003). For educational researchers, this “breadth over depth” pedagogical approach is highly questionable (Levstik, 2000; Pahl, 2003a; Rees, 2003).

To professionals in the field of history, the “breadth over depth” approach is just one of the many ways that standardized tests pervert the process of history (i.e. how historians actually “do” history). As mentioned previously in the section on conceptual literature on HSS standards, historians are particularly wary of the effect that standards and assessments have on students’ perception of the process of historical inquiry. Rothstein worried that testing “inevitably creates incentives for teachers to teach history as a succession of relatively meaningless facts” (p 1390). Wronski (1993) argued this approach may be easily quantifiable but it is not representative of the profession of history. Indeed, he argued, the process of historical study requires making value judgments; the very nature of history is “not always amenable to quantification” (p. 21). Wineburg (2004) also noted that of the five aspects of historical sense, the “ability to answer factual questions about historical personalities” was “the narrowest, and...the

least important type of historical ability (p. 1402). Strange, he muses, that this approach to history has come to define the curriculum.

Madaus' principle argued that standardized tests define the curriculum and content that HSS teachers employ. He also argued that tests define the assessments that teachers use in their classrooms. Namely, scholars argue that state-mandated tests increase the prevalence of multiple-choice tests in the classrooms. Kurfman (1991) claimed that teachers' increasing use of multiple-choice tests represents their desire to improve the "test-wiseness" of their pupils (p. 318). The logic is that, by getting the students used to multiple choice tests, teachers can help improve their students' score. Kurfman noted one empirical study (Mathison, 1988) that found teachers used multiple-choice tests as a way of "providing students with the experiences relevant to the standardized format" (p. 318), yet, this finding was limited to this one study. Kurfman concluded that the assumption in the conceptual literature that teachers will change the format of their classroom assessments based on standardized tests is yet another area in need of additional empirical research. With or without the empirical research to base their claims, scholars time and again point out that while quality teachers make changes to their curriculum and instruction to improve students' academic skills and knowledge, tests pervert that relationship. Testing, Volger (2003) claimed, promotes an environment wherein changes to curriculum and instruction are designed not to improve students' learning, but to improve students' test scores (p. 208).

Even if it were the case that teachers were administering multiple-choice tests in their classrooms to prepare their pupils for standardized tests, what harm can come from

it? The conceptual literature is rife with assertions regarding the bad pedagogy of multiple-choice tests. Pahl (2003c) claimed that pupils do not retain a great deal of historical information when factual tests are the predominant assessment used in a classroom. Further, when tests focus on memorization skills rather than on historical inquiry skills required to gather, examine, analyze, and apply information, the tests cheat pupils of the skills they will need in the future (Pahl, 2003a). Kurfman (1991) claimed that multiple-choice exams are likely to have “serious consequences” for the amount of actual writing that teachers ask students to do (p. 314) leaving a generation without practice and experience with conventional writing structures. Finally, Kurfman argued that students who think in more basic terms may see a rise in their scores when multiple-choice tests are the norm. This occurs to the detriment of high-level thinking students for whom results on a fact-recall test do not represent accurately their abilities.

Not all standardized tests are low-order, multiple-choice tests, however. In fact, Volger (2003) argued that standards based reform allows and welcomes varied methods to teach to standards and varied assessments to check that standards are being met. Horn (2006) asserted that “well-crafted, well-implemented, well-used assessments can be developed” (p. 58). Yet the difficulty of devising these measures and devising multiple-choice items that measure higher-order cognitive skills is well noted in the literature (Hirsch, 1996; Kurfman, 1991; Pahl, 2003a). Higher-order tests are often rejected in favor of low-order tests because high-order tests are more expensive and time consuming, and less reliable both to write and to score (Pahl, 2003a). “Everyone agrees,” Rothstein (2004) charged, “that students need a fact base to make critical decisions but you can’t

test the latter so we only test the former “ (1389). Madaus’ proposition that tests alter both the subject matter taught, and the assessment measures used in the classroom continues to reverberate in the conceptual literature on history and social studies assessments.

Principle 6: When test results are the sole or even partial arbiter of future educational or life choices, society tends to treat test results as the major goal of schooling rather than as a useful but fallible indicator of achievement.

As eluded to in the background section of this paper, the attention given to the performance of pupils on tests – particularly by school teachers, administrators, education professionals, and media pundits – leads many in the public to assume that test scores accurately represent a meaningful level of achievement for the student taking the test. While the true purpose of schooling is hotly debated in academic circles, history and social studies teachers often contend that the task of their discipline is to ready pupils for democratic citizenship. Do social studies tests accurately measure a pupil’s readiness for this responsibility?

Most testing critics answer this question in the negative. They claim that those who favor HSS tests fail to consider the content validity of the test – that is to say, whether the test actually measures a student’s aptitude to perform the tasks, skills and logical thinking required of citizenship (Horn, 2006; Kurfman, 1991; Pahl, 2003a; Rothstein, 2004). Validity or “content validity” specifically, Pahl (2003a) reminded us, refers to a test’s ability to ask questions about a topic that experts in the field agree is important” (p. 212). Validity also refers to an assessment’s ability to measure accurately

the knowledge base of the students taking the test” (Pahl, 2003a, 212); or, as Horn put it, “does content domain taught match content domain measured?” (Horn, 2006, 65). Rather than validly measuring a pupil’s readiness for citizenship, testing critics claim the tests measure something else entirely.

Some claim that test-writers construct items that focus on meaningless trivia. Students who succeed on these measures, therefore, are students who display an ability to memorize random, perhaps unimportant, information. While some historical information is arguably more important than other information, relevant and irrelevant information is treated with equal weight. Rees described two items on a standardized history test to exemplify this point; one item asked students about the origination and meaning of the quote, “We hold these truths to be self-evident...” A second item on the same test asked students to identify information related to the invention of the steel plow. Despite what Rees saw as the clearly variable degree of importance of these items, they were scored equally. Pahl (2003c) questioned this practice and maintained that much of the content included in factual history tests is “obsolete.” When facts are held to be the chief aim of history class, he claimed, history becomes a subject with “zero utility” (p. 199). What can be gathered then, about a student who is successful on such an examination?

According to most scholars, very little information can be gleaned about students’ ability from test scores (Horn, 2006; Pahl, 2003a, 2003c; Rees, 2003; Rothstein, 2004). Certainly, they argue, the test scores say little about the students’ likelihood of performing the “rights and responsibilities of citizenship,” (Rothstein, 2004, 1385). In fact, a *high* test score can sometimes mean the student is *less likely* to act civically.

Rothstein (2004) pointed to studies that show that African-American students, who traditionally score lower than White students, are more likely to engage in civic activities. Meanwhile, research shows that particular pedagogical activities (e.g. debate, discussion, service learning) are correlated highly with civic participation (Rothstein, 2004), no research has shown that acquisition of historical facts alone leads to increased civic activity (Rothstein, 2004). Indeed, Pahl (2003c) questioned whether we have sufficient evidence to claim that teaching and testing history actually trains students to make democratic judgments, creates sound habits of study or broadens the students' point of view (p. 199).

Madaus' principle, then, appears to hold in the conceptual literature on HSS assessments. While scholars note that the public places a great deal of emphasis on the result of the tests to determine a student's ability, scholars argue that little information can validly be gleaned from the pupil's score on the assessment. Bad test scores in history, Gaudelli (2002) claimed, does not necessarily show that students "know nothing about history," rather, it's more accurate to say that they simply can't recall facts. If the public wants educators to make high stakes decisions about what history students should know and whether or not they have been prepared for citizenship in a democracy, scholars argue they should look to multiple data points – not just a standardized test in history and social studies – before they make those decisions (Pahl, 2003c).

Principle 7: The more any quantitative social indicator is used for social decision making, the more likely it will be to distort and corrupt the social processes it is intended to measure.

If standardized assessments in history and social studies are meant to measure learning in those topics, Madaus would claim that the tests will distort and corrupt the learning process. To be sure, if the “learning process” includes everything from introduction of key concepts and skills, instructional approaches, accurate evaluation of students’ learning in the discipline and adjustment of pedagogical approaches to address learning needs, Madaus’ principle, when applied to HSS, appears to be prophetic yet again. Academics argue the tests distort the learning process by sorting students into ability categories rather than focusing on an individual’s learning needs. And, rather than using student’s rankings to improve learning, results are used against the students’ learning interests. First, I address the literature that claims the standardized assessments sort students. Next, I discuss literature that describes how the scores and rankings are used in cross-purposes to the learning goals of many teachers.

Standardized test makers use many techniques to sort students. As established in previous sections, standardized assessments in HSS do not necessarily test students on constructs that teachers deem central to the discipline. Oftentimes, tests focus on mundane, random, even meaningless data. Why does this happen? Wouldn’t it benefit teachers to assess students on content deemed essential? Test-writers would claim that a test that focuses on basic content would not produce the results that test-makers would

like see. That is, the test would not produce what psychometric experts call “spread” (Pahl, 2003a; Savage, 2003; Wineburg, 2004).

The purpose of a good test, as test-writers would have it, is to accurately discriminate between students who are top performers and students who are poor performers. A useable item on a standardized test, therefore, is an item that “good students always answer correctly, and that poor students always miss” (Pahl, 2003a, 213). While many in the general public do not differentiate between a norm-referenced and a criterion-referenced test (Savage, 2003; Volger, 2003), the distinction is critical for test makers. With criterion-referenced tests, all students reasonably could answer all of the questions correctly. In a norm-referenced test, test-makers chose items based on spread. They look for items that a predictable group will answer incorrectly – regardless of how critical or trivial the historical subject matter is.

If test-makers are constructing a norm-referenced history test, important facts and trivial data are often scored equally. What results is the very perverted process that Madaus described in his principle. Wineberg (2004) wrote,

Facts do not appear because of inherent worth or because they were taught in the high school curriculum or even because a blue-ribbon commission declared that every American high school students should know them. Rather...countless bits of information appear on the test because they work mathematically; they snare their targets in sufficient numbers to boost the item’s discrimination index (p. 1408).

If too many students are able to answer too many test items, test-writers introduce “distracters,” or items that are likely to sort the students along a continuum (Savage, 2003; Wineburg, 2004).

A critical eye toward the results of the sorting process would quickly discern yet another example of the tests’ power to distort the learning process. Upon examining how the students sort or “spread,” one may notice what Pahl described as a nearly perfect correlation between high-SES schools and high scores on the one hand, and low SES schools and low scores on the other (Pahl, 2003a). Indeed, Popham (in Savage, 2003) found that about 45% of social studies items were linked to socioeconomic status (p. 65). Yet, as Pahl (2003a) contended, this near-perfect correlation between scores and SES is virtually ignored in most of the literature on the outcomes of standardized assessments in history. Based on the test results it would appear to most that students from lower socioeconomic classes are unable to learn history and social studies. Yet, the poorer performance of low SES students on the standardized history tests is more likely an indication of problems outside of the learning process, and therefore outside the control of both students and teachers. Test results, therefore, are not a reflection of a students’ capacity or aptitude for learning history and the social studies but rather, as Volger (2003) put it, test results have “almost everything to do with parental socioeconomic backgrounds” (p. 208).

Still others argue that test scores are neither a reflection of students’ aptitude or capacity, nor a reflection of parents’ socioeconomic background. Rather, tests results are a reflection of, even a ranking of, a school district’s teaching resources including teacher

capabilities. Savage (2003) highlighted the fact that teachers at schools with high socio-economic standing are more likely to be evaluated as more effective than those at low SES schools (p. 203). Would it only follow that students' test scores are not a reflection of their ability to learn but of their teachers' ability to teach? If it is indeed the case that standardized test scores reflect parents' socio-economic class, teachers' (in)ability to employ effective pedagogy, and schools' resources along with students' historic literacy, Savage argued that placing emphasis on the results of these tests throws into question "the basic principles of democracy, which focus on the development and the well-being of all citizens, especially those who are most fragile" (Savage, 2003).

Another way that the conceptual literature claims the learning process is corrupted by standardized tests is in the way the test results are used to inform and/or reform instruction. First, Pahl (2003c) noted that rather than being formative tests (wherein teachers review results of the test and adjust their teaching accordingly) most standardized tests are summative. Tests that are summative usually occur near the end of the year and offer teachers little to no ability to address learning deficiencies in their students and help students improve their test scores.

When a school does poorly on a standardized test and teachers have an opportunity to address the learning needs of their students, an incentive is created for schools to take short-cuts to improve students' scores. That is, rather than addressing the learning needs of the entire population of that school, administrators focus on those whom have come to be called the "bubble kids." The "bubble kids" phenomenon has been discussed broadly in education research literature (Booher-Jennings, 2005;

Kurfman, 1991). Savage (2003) described them as the students “most likely to change with the least amount of effort” or “those at or just below the mean” (p. 203). When administrators and teachers focus on raising the scores of the bubble kids, so that the mean score for the school will subsequently rise, the learning needs of those students at both extremes of learning success – the extremely low achieving and the extremely high-achieving – will be largely left out of the educative process (Savage, 2003, 203).

Scholars argue that test scores are so sensitive to social phenomena having nothing to do with actual HSS learning that relatively significant increases and drops in school districts’ scores on standardized tests have little to do with the quality of teaching in a school, or even to do with increases and decreases in student learning. Savage conceded that “a simple demographic switch of a few children from different socioeconomic groups can significantly alter the average scores” (p. 202). The fact that a district’s scores can swing significantly from year to year based on a few children, rather than changes in content delivery in the school, further calls into question what scholars describe as the arbitrary nature of cut-scores (Horn, 2006). When the addition, or removal, of just a handful of pupils in a school district can mean the difference between being labeled “proficient” and “in need of improvement,” one must wonder whether the cut scores accurately reflect teaching and learning quality found in the school (as they are intended to do) or if they are a reflection of minor demographic shifts in the population. Certainly, HSS scholars have noted throughout the literature the tendency of tests to distort the social process that they are meant to measure. Standardized tests have been

transformed, they argue, from instruments intended to measure the learning process into instruments intended to sort and process.

Principle 1: The power of tests and examinations to affect individuals, institutions, curriculum or instruction is a perceptual phenomenon: If students, teachers, or administrators believe that the results of an examination are important, it matters very little whether this is really true or false – the effect is produced by what individuals perceive to be the case.

Calls for reform and testing in education tend to peak and recede in correlation with how the public perceives things are going, generally-speaking, for the country. The sense that the call for tests and examinations, and the effect of tests and examinations, is a perceptual phenomenon is widely echoed in the conceptual literature on HSS assessments. Kornblith and Lasser (2004) wrote that tests are called for when taxpayers wonder if they're getting their money's worth in the campaign to raise educational standards. Politicians respond to this call, Pahl (2003b) claimed, by generating interest in testing. Scholars have noted this trend of vacillating support for tests and point to shifting state policies as evidence. One scholar (Grant, 2006) asserted that tracking states with high-stakes, mandated tests in history and social studies is like watching a moving target. From year to year, the numbers of states with tests ebb and flow in concert with the whims of politicians' and their constituents. Ironically, not just tests but test *results* contribute to this perceptual phenomenon. When test results show dismal student performance, they predictably fan the flames of public unease, even discontent, about

education. Test results, therefore, perpetuate and intensify calls for increased testing and accountability.

Several scholars in the field scoff at the public's (and public official's) misdirected calls for testing and their subsequent reactions of disgust with the results (Rees, 2003; Rothstein, 2004; Wineburg, 2004). These scholars note the tone of politicians who rally their constituents with demands that education professionals "return" to the basics. They hearken back to the day when children knew the important facts and dates, when a high school graduate could distinguish the Revolutionary War from the Civil War; the Declaration of Independence from the Gettysburg Address. Scholars in the field of history and social studies unanimously respond that that Golden-era of historic literacy never occurred (Feldman, 2007; Rees, 2003; Rothstein, 2004; Wineburg, 2004). Much like Socrates' observation that children of his day were more tyrannical, and therefore inferior to children who came before them, scholars claim that contemporary education critics are imagining the declining state of children's natural abilities. It is simply a matter of false perception manifesting as a reality. In the meantime, as Rothstein asserted, "no standardized, norm-referenced test will show kids as historically literate – 50% will always do poorer than average! (1409)." That is the very nature of a norm-referenced test.

Meanwhile, despite the dubious incentives for, and outcomes of, standards-based, state-mandated history and social studies tests, the public is often suspicious of teachers who balk at calls for increased testing. Academics note the tendency of public officials to lay blame at the feet of teachers who won't consent to tests. They claim that these

teachers don't want to be held to standards and are shirking their responsibility to children. Proponents of high stakes assessments, in particular, contend that opposition to testing "is rooted in the desire of educators to avoid accountability and maintain the status quo" (Savage, 2003, 2). Regardless of whether tests drive the curriculum, generate intended results, accurately represent students' achievement, or validly measure important constructs, the public perceives tests to be doing all of the above. And, perception, Madaus claimed, has become reality. In previous pages I have noted that authors of conceptual pieces assert that many of Madaus claims have risen to the level of truism. However, they note, empirical studies are needed to verify that these principles can be found in practice. In the following section, I review the empirical literature on history and social studies standards and assessments to see if and how the phenomena Madaus describes has surfaced in history and social studies classrooms.

Empirical studies on standards and assessments

The literature search unearthed 27 empirical articles and reports on standards and assessments. These studies were not divisible between articles that focused on standards and articles that focused on assessments. Generally, in the empirical studies, articles that focused on assessments also discussed standards in an in-depth manner. The reverse was also true – the few empirical studies of state and national standards tended to discuss assessments. As stated, empirical articles that addressed the impact of standardized testing across all curriculum areas were weeded out save three studies often cited in the HSS empirical literature (Cimbricz, 2002; Grant, 2000; Pedulla, et al., 2003).

Additionally, empirical articles that focused on the impact of HSS subject-specific standards and assessments on students were also discarded.

Empirical Studies That Legitimize the Call for Standards and Assessments

Most empirical studies that have anything to say about when the HSS testing mania began point to two sources – the national government’s 1983 report, *A Nation at Risk*, and Diane Ravitch and Chester Finn’s 1987 book, *What do our 17-year-olds know?* The *Nation at Risk* report was based on the findings of the National Commission on Excellence in Education. Based on papers presented by members of the Commission and the testimony of “administrators, teachers, students, representatives of professional and public groups, parents, business leaders, public officials, and scholars” the Commission found that the “declines in educational performance are in large part the result of disturbing inadequacies in the way the educational process itself is often conducted” (National Commission on Excellence in Education, 1983). In a now-famous quip, the *Nation at Risk* report concluded, “If an unfriendly foreign power had attempted to impose on America the mediocre educational performance that exists today, we might well have viewed it as an act of war” (National Commission on Excellence in Education, 1983). The report spurred debates at the national level as to how the country can reverse the supposed decline in educational excellence in our nation’s schools.

Meanwhile, Ravitch and Finn’s study was prompted by the work done for the Educational Excellence Network, a group which, in the years between 1981 and 1986, became increasingly concerned with “the aimlessness and intellectual torpor that seemed to handicap the [humanities] fields” (2). The authors’ empirical study was based on

findings from the “First National Assessment of History and Literacy” – a multiple choice test developed by two committees of content area professionals that consisted of 262 cognitive or “knowledge” questions – 141 in history and 121 in literature. The assessment was administered to a national sample of 7,812 seventeen- year-olds in public, private and religious schools in the spring of 1986. The sample was evenly divided between young men and women and the results were statistically weighted to proportionally represent the demographic makeup of American youth. The results of the assessment were, “little short of appalling” (p. 40). In fact, in the preliminary administration of the test, the authors described the overall performance of students on this history portion as “unsatisfactory” (p. 56), they described students’ knowledge of constitutional precepts as “shaky” and the Civil War as “ill understood” (p. 57). While there is little room here to fully disclose the results of the assessment, suffice it to say the results led the authors to conclude that “something is gravely awry: our eleventh graders as a whole are ignorant of much of what they should know” (p. 200).

These two reports laid the groundwork for much of the subsequent research on the historical literacy of America’s youth. Since the publication of the reports, three other empirical studies have since surfaced that either dispute or verify the findings of the *Nation at Risk* and Ravitch and Finn’s study. Whittington’s (1991) study called into question what he considered to be the main finding of Ravitch and Finn’s study – that is to say, he challenged the fact that pupils’ knowledge of canonical history and literature had declined sharply in the decades leading up to the report.^{vii} Whittington’s comparative study matched test items from large-scale history test administrations in 1917, 1933,

1944, 1964 and the College Board Achievement Test in American History from 1982 to Ravitch and Finn's test administered in 1983. The author matched items based on subject matter and the items' level of difficulty. By comparing student performances on questions about the same topic and about the same difficulty, the author found little evidence to substantiate what he viewed as Ravitch and Finn's insinuation that the historical knowledge of America's youth was indeed in decline. Rather, students in the 1980s, he wrote, are "not demonstrably different from students in their parents' or grandparents' generation in terms of their knowledge of American history" (Whittington, 1991, 776).

Hess' (2008) empirical study was a replication study using Ravitch and Finn's original "National Assessment" instrument. His study found that little had changed in the 20 years since the original publication. Instead, the report documented "continuing weaknesses in our students' knowledge of history and literature" – a fact that contributors to the report contended was a direct result of calling for "basic skills" rather than demanding "excellence in education" (p. 3). The author doubted that students' can be adequately prepared to participate in a democracy without knowing the "historical narrative and cultural touchstones that mark our national experience" (p. 5). Hess hesitated to claim a "decline" in historic literacy. Rather, that author noted that the level of achievement of American students is a continuation of the poor performance uncovered by the 1987 study. It may be deduced that the author would concede that the golden-era of historic literacy (discussed earlier) is a fallacy.

A final empirical piece found deep concern among a sample of high school teachers about the civic future of the United States. Feldman's (2007) exploratory study conducted in 2006 investigated the perceived "state of civics education" as seen by practicing teachers in three large high schools (p. 7). The three schools were a purposive sample – the researcher had personal familiarity with each. The schools enrolled between 2,200 and 2,700 students. Surveys were mailed to all full-time teachers in every subject area in the three schools. On the survey, teachers were asked to respond to three Likert Scale questions regarding their perceptions of their own civic preparation and the civic preparation of their students. The researcher found the greatest level of variance in teachers' responses to the question about their own civic preparation – that is, some teachers felt civically prepared and some did not. Much of this variance could be explained by the number of years of teaching experience the participant had. The results, however, showed much more consistency with regards to teachers' perceptions about the civic preparation of their students. While the author conceded that the sample was not representative of the national population of high school teachers, he noted that it was apparent that a consensus existed between teachers from the three high schools – they were all deeply concerned about "the lack of knowledge, attitude and participation of young people necessary for the future growth of the American representative republic" (p. 11).

Empirical Studies That Analyze the Ebb and Flow of Standards and Assessments

The alarming findings presented in *A Nation at Risk, What do our 17-year-olds know?* and successive studies led to a call for national standards and assessments. The

reports also led to demands for more rigorous teacher certification and licensure standards. Several empirical studies chronicle the changes in HSS standards and assessments following the two reports.

Henry's (1987) content analysis of the Advanced Placement American History examination from the years 1960 – 1984 found a marked shift in the way that exam assessed historical and social scientific knowledge. Most notably, the study found a significant decrease in the essay requirements of the test and an increase in the emphasis on multiple choice questions. His analysis of 2,035 multiple-choice questions on the test also found a significant decline in the number of questions that utilized primary documents and questions that required the interpretation of symbolic documents. Meanwhile, questions that did not call for students' response to stimulus materials (i.e. fact recall questions), questions that focused on traditional social science, arts and humanities material increased significantly. While this study did not mention *A Nation at Risk* specifically, the author does conjecture that the changes in the AP American history exam were precipitated by the testing board's concern that the education reforms of the 1960s and 1970s – featuring inquiry use of primary documents and symbolic representations materials – were “not appropriate” on the advanced placement American history examination (p. 161). The committee's change in stance may have been brought about by the *Nation at Risk* report or, if not by the report itself, may reflect shifting trends in thought about the appropriateness and effectiveness of inquiry methods in history and the social studies.

Four empirical studies chronicled the changes in state standards and assessment practices in the years following the *Nation at Risk* report. Lobes (1998) empirical piece was based on a study conducted by the National History Education Network (NHEN). In 1997, the NHEN collected all available state standards for history and social studies. Using an undisclosed methodological approach, the NHEN found that state standards generally echoed the standards frameworks of one of three organizational structures; the National Council for the Social Studies (NCSS), the social science model (SS) and this history-social science model (HSS). Lobes study found that 31 states (the vast majority of states with written standards on history and social studies) followed closely the NCSS standards model. Six states followed the second organizational structure (SS), and eight states followed the third organizational structure (HSS) in the creation of their standards. In the final HSS model, history served “to integrate the other social sciences” (p. 230). The author pointed to the small ratio of states using the discipline of history to ground the social science discipline, and concluded that professional historians must work harder to make their professional interests’ a greater priority in the development of state standards and assessments in history and social studies.

Buckles, Schug and Watts (2001) survey of state standards and assessments was based on telephone interviews conducted in 1999 with state department of education social studies representatives or assessment supervisors. Their study began by outlining the five sets of national standards published for the social studies and other subject areas taught by social studies teachers. The five sets of standards included standards written by the National Council for the Social Studies, the Center for Civic Education, the National

Center for History in the Schools, the Geography Education Standards Project and the National Council on Economic Education. Buckles, Schug and Watts study found that forty-four states had set standards in social studies, and thirty-four states conducted annual assessments of student achievement in social studies.

The findings of Buckles *et al.* study are interesting in light of the findings outlined in *Education Week's* annual *Quality Counts* report (Swanson, 2008). The *Quality Counts* report is similar to the Buckles *et al.* report in that both reports surveyed the state standards and assessment practices in history and social studies. Yet their findings are vastly different. The 2008 *Quality Counts* report found that 20 states have “clear, specific and grounded standards” in history and social studies. This number is nowhere near Buckles’ *et al.* 2001 finding that 44 states had standards in social studies. The *Quality Counts* report also indicates that only 20 states have high school history and social studies assessments that aligned with state standards. Again, this number is significantly different from Buckles’ *et al.* finding that 48 states are administering tests in social studies. Though the *Quality Counts* report makes clear that many states have a long way to go in the development of their standards and assessment practices in the subject of history and social studies, the report says little about the apparent decrease in the number of states that have standards and assessments in history and social studies since the time of Buckles’ study. Perhaps the significantly different findings lend validity to Grant’s observation (discussed earlier) that tracking a state’s standards and assessment practices is like watching a moving target. The difference in the findings may simply suggest

dissimilar methodologies in determining what counts as state standards, and what counts as a state assessment in history and social studies.

A final study on the shifting patterns of state standards and assessments focused not on standards and assessments for students, but for history and social studies teachers. Brown's (2006) investigation of teacher certification in the states zeroed in on content standards for teachers. In a massive data collection and organization effort, Brown and her colleagues gathered each state's online documents on teacher certification in history and social studies. They then organized the state teacher certification standards into categories that included guidelines and rules for certification, and content knowledge required for certification. Brown found significant state-to-state differences in certification practices centered on four categorical differences including: a) the number of required hours of study in a discipline or field of study, b) demonstrated proficiency in the subject, c) college and university requirements for certification programs, and d) levels of difficulty on state administered content tests. Brown asserted that the intrastate variance in certification standards highlights a shortcoming in the historical and social science disciplines. She charged historians to create and adhere to strict standards and to "set their house in order if they do not wish it to be remodeled without their consent" (p. 379).

Empirical Studies That Critique the Standards and Assessments

Two final empirical studies address the continued concern over students' civic knowledge and skills, and the subsequent calls for standards and assessments in history and social studies. However, both studies are deeply critical of the manner in which

standards and test makers have approached their mandate to improve civic education. Stern (1994) conducted a content analysis of the *National Standards for United States History* compiled by the National Center for History in the Schools and released in 1994.^{viii} Stern analyzed the history standards based upon three criteria: 1) Do the standards equip young Americans with the skills for mastering historical thinking? 2) Do the standards help students grasp context and develop a sense of history? and 3) Do the standards replace the myths and distortions of conventional history with a more truthful and balanced perspective of the past (Stern, 1994, 62)? Based on these criteria, the author judged the standards to be woefully inadequate. He claimed that the standards led students into a “trap of presentism” – that is, they ask students to judge the actions of historical actors based upon contemporary norms and values. Secondly, the standards fail to tell an accurate and balanced story about the American past. Rather, they omit significant details such as the role of Black Africans in the slave trade. The author concludes that the history standards fail to encourage in students an accurate, useful and balanced sense of history.

A second empirical study of state standards and assessment practices that is deeply cynical about standards and tests is Kelly, Meuwissen & VanSledright’s (2007) piece, *What of History?* Using Seixas’ framework that delineates three approaches to history (the collective memory approach, the disciplinary approach and the postmodern approach), the authors conducted a content analysis of the *National Standards for History* published by the National Center for History in the Schools, the state of Virginia’s *History and Social Science Standards of Learning for the Virginia Public Schools*, and

Fairfax County Public School's curricular materials to "examine the conceptions of history forwarded by these policy instruments" (p. 118). The authors found that the standards and testing movement legitimized only certain types of historical study, concepts and interpretations, and delegitimized others. Specifically, the national, state and district standards emphasized collective memory and historical content over historical inquiry and thinking skills. The prioritization of content over skills, the authors contend, encourage teachers to adopt "pedagogical moves" that run contrary to the research on best practices in social studies teaching and learning (p. 119). "The subtext suggests," they write, "the subordination of thinking to knowing and recalling" (p. 128). When standards use this line of attack to deal with the supposed decline in historical literacy of America's youth, the authors contend, "common practices with all-too-typical results will almost certainly prevail" (p. 138).

Empirical Studies on the Effect of Standards and Assessments on HSS Teachers' Practice

The vast majority of conceptual articles on standards and assessments claim that they will significantly affect teachers' curricular and pedagogical choices. However, many of those same articles note that there is little empirical evidence to demonstrate that this assumption is true in reality. Here, I review the empirical research on the effect of standards and assessments on teachers' practice. First, I address three studies that analyzed the effect of standards and assessments on the practices of teachers across disciplinary content. Second, I address empirical studies of history and social studies teachers' *perceptions* of the impact of standards and assessments in their classroom. In the final subsection, I review the empirical literature on the impact of standards and

assessments in history and social studies classrooms as conveyed through classroom based, small-scale, qualitative studies. As I make clear in the concluding thoughts, the qualitative studies supplement the larger-scale quantitative surveys of teachers' perceptions in significant ways.

Effects of Standards and Assessments on Teachers' Practice Across Disciplinary Content

Dozens, if not hundreds, of empirical studies exist that explore the impact of standardized testing on the school curriculum. These include studies of standardized tests on students, teachers, administrators and the public. I include these three articles in the review, and discard the dozens of others, for simple reasons. One article is a review of the empirical research on the impacts of standards and assessments *on teachers practice*. One article is a study of teachers' perceptions about the impact of standards and assessments that is frequently cited in the literature. The final study was conducted by a prominent scholar in the field of history and social studies – his study resulted from, and resulted in, many more studies specifically to do with the field of history and social studies.

Cimbricz's (2002) review of the empirical literature on the relationship between testing and teaching found that most empirical studies claim state mandated testing has either a negative or a neutral effect on teachers. A very small minority claimed that state mandated testing had no effect or a positive impact. Of the four studies that showed state mandated testing had a negative effect on teachers practice, each claimed the tests impacted teaching in five key ways. They claimed that tests: a) narrow the curriculum, b) foster anxiety, confusion and shame among teachers, c) deskill teachers, d) are an invalid

measure of concepts taught, and e) rob teachers of instructional time forcing them to focus on test preparation and administration rather than disciplinary content. Five studies showed that state mandated tests had an overstated effect on teaching. These studies proclaimed that the influence of tests was especially weak when it came to teachers' choice of pedagogical approach. In other words, state mandated tests influence the content of instruction but they did not influence how teachers teach. Her review also found that studies suggested the influence on teachers was qualitatively different at differing grade levels and years of practice. The author concluded that very few empirical studies exist that discuss the relationship between standardized tests and teaching. The author called on the professional field to conduct additional empirical studies to describe and clarify this relationship.

Abrams, Pedulla and Madaus' (2003; 2003) review of the literature and massive empirical study of teachers' perceptions of the effect of state-mandated testing programs found that tests impact teaching practices in concrete and negative ways. The authors administered their 80-item survey to 12,000 teachers in states with varying levels of stakes attached to the state-mandated exam. 4,195 teachers returned useable surveys – a response rate of 35%. The authors found that regardless of whether teachers were from states with high stakes tests or low stakes tests, the presence of the test alone led them to “teach in ways that contradict their own notions of sound educational practice” (p. 23). A significantly larger number of teachers from states with high-stakes tests claimed they felt pressure to raise their test score than did teachers from states with moderate to low stakes tests. These teachers also felt that they had little time to teach anything that was not

emphasized on the test. The authors concluded that state tests, rather than the state content standards, are the “more powerful influence on teaching practices” (p. 27). They called on state testing programs to expand the role of, and solicit additional guidance from classroom teachers in the construction of state mandated assessments.

The final study on teachers’ perceptions of the impact of state mandated tests across the disciplines came from S.G. Grant, a prominent scholar in the field of history and social studies assessments. Grant’s (2000) exploratory study using focus group interview data collected over two years from cross-subject matter groups of elementary and secondary New York state teachers presented, what Grant called, a “mixed picture” of the impact of tests on teachers. Grant used the proxy “reform by rumor” to explain how the New York state high-stakes Regents test impacts teachers at the classroom level. That is to say, teachers’ perceptions of the test are predominately informed by casual discussions with other teachers. In these conversations, teachers both attempt to make sense of the test and predict possible consequences of the test. Grant’s study found that teachers were uncertain about the tests. Many believed that the tests would be used to blame teachers or that teachers would be used as scapegoats for the poor performance of their pupils. Teachers also predicted that the tests would be ineffective at leveraging educational reform. While some teachers predicted negative consequences for their pedagogical practices (including reductionist teaching strategies like lecture and “drill sessions”), others assumed that the test would have little consequence for their teaching. Grant concluded by noting, “faith in tests as a means of corralling teachers’ practices may

ultimately prove chimerical” unless teachers are included in the decision making and construction of state mandated tests (p. 21).

The Perceived Impact of Standards and Assessments in HSS: Survey Research

The literature search turned up three empirical studies that analyzed history and social studies teachers’ perceptions of state-mandated standardized tests using broadly themed survey instruments. Burroughs (2002) nationwide survey of more than two hundred social studies educators found that a vast majority of social studies educators across grade levels (71%) indicated that “state-mandated content area testing has had an impact on curriculum in their classrooms.” In the open-ended responses to the surveys, teachers lamenting the negative repercussions of state-mandated testing outnumbered positive statements by 2:1. The author stated that, upon closer inspection of the data, “a more detailed and holistic image [arises] of what is happening in classrooms.” Yet, the author does not specify what she means by this. Additionally, it is unclear whether her definition of “state mandated testing” is limited to, inclusive of, or excludes entirely testing in history and social studies. The effect of state-mandated testing on history and social studies teachers would be vastly different, one would assume, for a teacher in a state with a mandated test in HSS versus a teacher in a state with no mandated HSS test. Despite this marked oversight, Burroughs concludes by urging social studies teachers to enter the debate over state mandated testing.

A few years later, Burroughs published another study based on data gathered from surveys, focus group interviews and individual interviews conducted with thirty-four elementary, middle and high school history and social studies teachers (Burroughs, et al.,

2005). The results of that study showed that in states with no mandated social studies assessment, instructional time for social studies had been reduced. Teachers were deeply concerned about two primary themes: First, teachers were concerned over the erosion of time devoted solely to social studies and second, the culture of the elementary school had shifted from a “greenhouse to a sterile environment” where, rather than learning, students were encouraged to practice test taking skills. Burroughs concluded that although teachers were distraught about how the *No Child Left Behind* Act prioritized English and math and diminished time for social studies, most teachers were reluctant to call for the inclusion of social studies in the high-stakes, state-mandated testing fray.

The final empirical study that surveyed teachers’ perceptions of state-mandated tests focused specifically, and explicitly, on the effect of state-mandated tests in history and the social studies. Volger’s (2006) study was based on survey data gathered from 107 Mississippi high school teachers. The author found the vast majority of teachers in his sample using teacher-centered, traditional approaches to teaching history and social studies against the recommendations of best practices by the National Council of the Social Studies. Vogler conjectured that the results suggest that teachers are teaching to the test – that is, social studies teachers are aligning their pedagogical content delivery specifically to cover the most amount of material in the least amount of time. Vogler’s findings lend some credence to the notion that high stakes assessments in history and social studies will lead teachers to favor breadth over depth in their content delivery. But, the author tempered this finding with a slightly cynical twist. Vogler suggested that state-mandated tests may not be the influential lever of change they appear to be. Rather, state-

mandated tests in history and social studies may just provide a “convenient excuse for teachers to engage in less ambitious teaching” (p. 293) such as teacher-centered, transmission style approaches to teaching. It’s difficult to know whether the transmission orientation to teaching is a result of, or unrelated to state-mandated tests in history and social studies.

The Impact of Standards and Assessments in HSS: Research in Classrooms

The remaining empirical studies on the effects of state wide mandated tests on history and social studies teaching demonstrates that tests do impact teaching, but the impact is limited and varied (Grant, 2006). The empirical scholarship of S.G. Grant has provided significant evidence to contest the notion that tests drive curricular change in school. In three separate studies of teacher’s responses to state-mandated tests (specifically, New York’s Regents exam), Grant found that the effect of state-mandated exams on teachers’ curriculum was limited. Moving chronologically, Grant’s first study (1996) on the influences of social studies teachers’ thinking and practice was based on data gathered from observations and unstructured interviews with 11 elementary, middle and high school social studies teachers. Grant found several influences on teachers’ content and pedagogical decisions; influences that went well beyond policy decisions such as state standards and assessments. Indeed, only one of the eleven teachers interviewed made mention of national reform efforts. Grant concluded that influences on teachers’ content and pedagogical decisions come from one of three levels: personal, organization (i.e. school or district sources) and educational policy. The three are

interactional such that influences that strongly impact one teacher may only modestly impact another.

Grant's second study (2001) on the influence of state-level testing on social studies teaching utilized data gathered from case studies of two high school social studies teachers in New York. Grant noted the prevalent assumption that state tests drive curriculum and pedagogy. Once again, his study did not provide conclusive evidence to support this assertion. Grants' case studies were based on two teachers who, according to one participant, were the "total opposite" of one another (p. 400). Grant found that one teacher, George Blair, employed a transmission-style orientation to HSS teaching. This approach to teaching seemed to benefit Blair when it came to preparing his pupil's for the state test – Grant asserted that Blair never mentioned the test, but his narrative instructional style meant he was preparing pupils for the test the whole time. Meanwhile, the second teacher, Linda Strait, was a self-described constructivist teacher. When it came to preparing students for the state test, this constructivist-oriented teacher had to work consciously and explicitly to prepare her pupils for the test. Grant concluded that the "faith in tests as a lever of change may be increasingly hard to sustain" (p. 422). In order for tests to clearly and consistently drive curricular change multiple factors must be in place – chief among those factors is the teacher's tendency to view the tests as a high priority and concern.

Grant's final study (Grant, 2007) was a summary of key findings and the literature review from his book project, *Measuring History*, which chronicled cases of state-level testing across the United States (Grant, 2006). He noted that while most conceptual

literature claims that teachers will drastically shift their content and pedagogical decisions to align to state requirements, the empirical data is not nearly as conclusive about that relationship. Some teachers, he wrote, react defensively and orient the vast majority of their teaching to reflect what they see as the requirements of standards and assessments. Still other teachers present examples of “ambitious” teaching. Grant echoed the assertion of an author in his anthology (Gradwell, 2006) by writing that these teachers teach ambitiously in spite of, not because of, state standards and assessments.

Three empirical studies focus on the impact of state standards and assessment on pre-service and beginning history and social studies teachers. Doppen (2006/2007) used a case study approach to analyze the reactions of pre-service teachers in his pre-service, undergraduate, social studies methods course to a section of Ravitch and Finn’s *Assessment of History and Literacy*. Doppen administered the test to his classes without notice and told his students that their grades on the assessment would be factored into their final grade in the class. Doppen claimed this approach would enable him to assess pre-service teachers’ beliefs about standardized testing and “identify some implications for how to better prepare them for their future role as classroom teachers” (p. 23). Doppen found that the participants in this study “struggled to understand the true impact” (p. 33) of standardized tests. While they questioned the adequacy of the test results to convey anything meaningful about a students’ learning in social studies, they were unable to critically analyze the content of the test. Participants in the study did not raise any questions about the multiple-choice format of the test, the test’s reliance on canonical history or the test’s general cultural bias. The author suggested that teacher educators

should better prepare pre-service teachers to work in a high-stakes environment. This entails preparing pre-service teachers to acquire the critical stance necessary to analyze the tests' appropriateness for an increasingly diverse student body.

The two other studies that researched the impact of standardized tests on beginning teachers found that while beginning teachers were concerned about their pupils' performance on the test, the tests did not impact beginning teachers' pedagogical decisions or lesson delivery. Van Hover used interview, observation and content analysis to analyze the impact of the Virginia SOLS on seven novice secondary history teachers. While the new teachers complained that the SOLS effectively outlined high priority content for teachers and dictated the scope and pace of content delivery, the author found that the SOLS had a weak impact in the novice teachers' pedagogical approaches. Similarly, Gradwell's case study of a third year middle-school social studies teacher in New York found that the novice teacher questioned the choice of content covered on the New York Regents Document-Based Question (DBQ) test. Nonetheless, the teacher never explicitly addressed the state-wide exam in her class. Grant suggested that this may be due to the fact that the DBQ aligned nicely with the teacher's pedagogical focus on documents. Interestingly, despite the documented differences the Virginia SOLS and the New York Regents tests, researchers in both studies observed beginning teachers employing both ambitious and pedestrian teaching approaches. None of the teachers appeared to intentionally adjust their instructional approaches to align with standardized tests.

The four remaining empirical studies analyzed the effect of state standards and assessments on secondary social studies and history teachers who had been in practice longer than three years. The findings of each of the four studies reflect Grant's assertion that standards and tests do influence teachers, but the effects must be qualified. State mandated standards and assessments seem to impact teachers in as many ways as there are teachers – that is, each teacher, each department and each district reacts differently to policy constraints. Fickles' (2006) case study of nine Kentucky teachers demonstrated how professionals react as a department to state policy instruments such as standardized tests.^{ix} This social studies department worked actively as a cohesive unit to align their department's content standards, pedagogical approaches and assessment practices with the state exam. The author qualified this finding by observing that teachers who taught untested subjects were not impacted by the state policies.

Yeager and Pinder's (2006) case study of two Mississippi teachers, meanwhile, found that the mere presence of a test, even if the test was not high-stakes, can influence teaching. Both teachers in the study made changes to their curriculum and pedagogy based on their understanding of departmental and district requirements. Yet, both teachers did what they could to make the standards and test work for them – meaning the teachers continued on their pre-test pedagogical trajectories. Salinas' (2006) interviews with eleven Texas teachers found that teachers may intentionally foster the impression that they are complying with standards and assessments when they actually have not in practice. Each of the teachers in her study claimed that they refused to turn over all content decisions to the state test – they surrendered content they viewed as unimportant,

and safeguarded the content they viewed as essential, whether or not it is likely to be on the test. While the teachers did concede to altering the pace of content coverage, they had learned to follow requirements set by administrators and state policy makers while remaining faithful to their teaching orientations (p. 189). These two studies support and reinforce one another. Both clearly show that teachers are aware of the standards but draw fine lines as to where they allow the tests to influence or direct their professional decision making.

Finally, Segal's interviews with five secondary social studies teachers in Michigan found that, perhaps due to the Michigan state test's emphasis on skills rather than content, the tests had very little impact on teachers' pedagogical decisions. The effect of the test on teachers, the author claimed, was ontological rather than pedagogical. The tests did not influence how teachers taught. They did, however, influence how teachers viewed themselves as teachers. Segal's study, therefore, lends credence to Kurfman's (1991) assertion that standards and assessments may have implications for teachers' sense of professional dignity and morale.

Concluding Thoughts on Standards and Assessments in History and Social Studies

In spite of assumptions and assertions made by most in the conceptual literature, the empirical literature seems to suggest that teachers are working with state standards and assessments in history and social studies. Teachers do believe that standards and assessments influence their teaching practices. However, standards and assessments are not, as suggested by some, a high-priority, or even the sole influence on their teaching practices. Some teachers continue to employ practices that many education researchers

would deem pedagogically unsound. These teachers chose to teach in this manner in spite of some states' recommended practices. Still other teachers employ ambitious pedagogy. These teachers also choose to teach in this manner in spite of state standards and assessments that appear to work against them.

There is, then, reason to hope that teachers, administrators and district officials have significant power – even in the face of state mandated standards and assessments – to practice as they please. Furthermore, they may also have some power to shape the discourse, direction and creation of these policy instruments. If mandated tests are here to stay, as many observers claim they are, classroom teachers and school administrators must press their cause with district and state policy makers. The dialogue between state policy makers, standards and assessment writers and classroom teachers must be open and meaningful if the conversation is to be the true democratic deliberation for which democratic pragmatists advocate.

The research reviewed here suggests that HSS teachers come from multiple and varied curriculum orientations. The research further suggests that teachers are reluctant to abandon those orientations – they often defy both professional training and policy instruments that work in cross purposes to these orientations. Establishing and encouraging true democratic dialogue between classroom teachers and policy actors will create the potential for state standards and assessments that can positively impact classroom practice without requiring teachers to turn their back on their deeply-held curricular orientations.

In the next chapter I begin with a discussion of Massachusetts high-stakes test in history and how it may implicitly be asking the vast majority of Massachusetts teachers to turn their back on their “deeply-held curricular orientations.” This will then lead back into the question, “What is the relationship between the aims of history and social studies teachers and state standards and assessments in history and social studies?” Then I introduce the survey methods and content analysis methods that I engage for this study.

CHAPTER THREE: RESEARCH DESIGN

The Massachusetts Education Reform Act of 1993 required state education officials to create curriculum standards for each of the major academic disciplines and to devise assessments to measure students' progression toward meeting those standards. History and Social Science is one of the academic subjects involved in this endeavor. Test writers in the state of Massachusetts have been working for over a decade to develop assessments which align with HSS standards in Massachusetts. But, while test writers assure the public that the assessment is aligned with state standards, no studies have been completed that analyze the relationship between assessments and teachers' orientations in the classroom. Given that findings of previous studies suggest that the curriculum teachers chose to enact in the classroom is a strong predictor of gains in student achievement (Gamoran et al., 1997), findings from this study could prove illuminating in the effort to boost student achievement in history and social studies.

For example, findings from this study may show that HSS teachers in Massachusetts claim to enact curriculum that aligns with state standards and assessments. Or, the findings may suggest that the reverse is true – that teachers report to enact curriculum that has no foundation in state standards and assessments. Misalignment between the two would suggest problems with either teachers' pedagogical and/or content decisions, or may call into question the content validity of the high-stakes, state-mandated standardized United States history assessment. Either way, the findings from this study will be of interest to state test developers, HSS teachers in high-stakes subjects

and education advocates concerned with alignment between standardized assessments and classroom teaching.

To that end, in this study I worked to answer three large questions:

1. What are the orientations of history and social studies teachers in the state of Massachusetts?
2. What skills and constructs are measured by the proposed high-stakes, state-mandated high school test in United States history?
3. What is the relationship between the orientations of history and social studies teachers in the state of Massachusetts to the constructs measured by the Massachusetts state-mandated history and social studies assessment?

To answer these three questions, I used a mixed-method approach which involved two separate data collection and analysis techniques. First, I used a survey with Likert-type items as well as open response items to collect data to answer the first question about HSS teachers' orientations in the field as well as their beliefs about standards and assessments. To answer the second question regarding the skills and constructs that the standardized assessments measure, I conducted a content analysis of a sample from the 10th and 11th grade MCAS assessment of United States History. To answer the third question about the relationship between teachers' orientations and the MCAS-US test, I conducted a Chi-Square goodness of fit test to compare the proportions of sampled Massachusetts HSS teachers who claim to work from each orientation to the proportions of test items on the MCAS-US test affiliated with each orientation. I also analyzed survey data that asked teachers about the Massachusetts HSS standards and standardized tests.

Design I: Survey of History/Social Studies Teachers

Data Collection

To explore the aims of teachers of high school history and social studies, I surveyed a random sample of 10th and 11th grade history and social studies teachers in the state of Massachusetts. In this section, I discuss the choice and modification of the survey instrument, the pilot study, the sampling procedure, the administration of the survey and the data analysis methods.

Instrument

Throughout the literature review process, I collected, reviewed and analyzed previously administered survey instruments from related studies of history and social studies teachers' orientations. Three empirical studies explored history and social studies teachers' orientations and employed survey methods. These studies also included reliability and validity analyses to verify the strength of their instruments (Anderson, et al., 1997; Barr, et al., 1977; Vinson, 1998). I have chosen one survey instrument in particular (Vinson, 1996, 1998) as a model for this study (See Appendix B). The construction of this instrument was based upon Martorella's theory of five orientations in the field of social studies and history. After a thorough review of the literature, I found that Vinson and Martorella's assumption of five orientations accurately reflected the classification schemes of a vast majority of the research in this field. Therefore, the Vinson instrument appeared to be best aligned to existing theory.

Vinson's (1998) empirical study was based on a survey of 490 high school social studies teachers and members of the NCSS and had a response rate of 45 percent. His

survey was comprised using components of Barr, Barth and Shermis' original Social Studies Preference Scale (SSPS) that measured teachers' endorsement of the philosophy, content and pedagogical approaches associated with each of the three traditions. Vinson added items to this survey to reflect Martorella's two additional categories: social criticism and personal development. He found that his research subjects identified more with the liberal leaning approaches of social studies as reflective inquiry (RI), social studies as informed social criticism and social studies as personal development than with the more conservative approaches, social studies as citizenship transmission (CT) and social studies as social science (SS).

The instrument was designed with six items targeting each of the five orientations. The six items were subdivided into three groups: two items reflected the purpose of history and social studies for that orientation, two items reflected the content choices preferred in each orientation, and two items reflected the preferred methodological approach of each orientation. Vinson used multiple items on the same instrument to get at the same construct to ascertain internal validity of the instrument. Vinson calculated within category correlations (Pearson's r) for each instructional approach's six corresponding items. The within category correlations revealed non-correlations to a moderate level of correlation between each of the six items meant to measure one of five instructional approaches. The lowest within-category correlation score was $-.0410$ $p = .830$ between items #10 and #25 (see Appendix B) in the Citizenship Transmission teaching approach. In other words, two of the six items that were meant to calculate a teacher's tendency toward the CT approach were not correlated with one

another. Meanwhile, the highest within-category correlation score was .71, $p < .0001$ between items #20 and #8 (see Appendix B) in the social science approach. In other words, two of the six items meant to calculate a teacher's tendency toward the social science approach were strongly correlated. The remaining within-category correlation scores ranged in strength from low to moderate (.3 to .4) with a few strongly positive correlations in each orientation.

The author postulated that many of the low to moderate correlations were due to several factors – most notably that one or two questions were double-barreled. For example, one question that Vinson theorized should appeal to teachers from the cultural transmission orientation read, “Lecture should be used as an instructional strategy because it enables students to acquire and retain a significant number of specific facts and concepts.” In order to agree with this question, a participant needed to agree with both the instructional method (lecture) as well as the rationale behind that instructional choice (to acquire and retain facts). Some of these double-barreled questions were revised before my administration of the survey. The intent was that this might boost the internal validity of the instrument.

A second theory regarding the low to moderate level of correlation for within category instructional approach is that participants' calculated instructional approach was based upon statements about the teacher's philosophy, pedagogy and curricular content choices. To have a moderate to high level of correlation for within category instructional approach items, participants had to make pedagogical and content choices that aligned with the philosophy of their instructional approach. Given Vinson's results, it appeared

that consistency between the three categories is not high. This finding exemplified the notion (discussed in the literature review) that teachers base their pedagogical and content knowledge decisions on more than just their own teaching philosophies. Additionally, teachers often mix and match one teaching orientation's content choices with another orientation's pedagogical or philosophical choices.

Vinson used a second method to test the internal validity of his survey instrument. One set of items on the survey represented a teacher's "calculated instructional approach." Participants answered six questions about five different possible teaching orientations. A participant's "calculated instruction approach" then was established as the highest mean score a participant had (i.e. the items with which the participant had the strongest agreement) for the six different teaching orientations. Next, participants were asked to select among five descriptions of the five different teaching approaches for the one that best represented their own teaching approach. This was the teachers "selected instructional approach." Vinson conducted a bivariate crosstabulation procedure (Chi-Squared) and a discriminant analysis procedure to analyze the relationship between calculated and selected instructional approach. Results of the discrimination analysis showed that the selected category accurately predicted their calculated category for 73.3% of participants in the study. Vinson wrote that this dependent relationship between calculated and selected instructional approach categories created "some level of confidence in the notion that both measured the same construct – that the questionnaire captured instructional approach" (p. 128-9).

Vinson discussed the ability to generalize the findings of his study to a larger population in his discussion of his sampling procedures. Vinson's sampling procedure resulted in a nationwide random sample of members of the professional organization of social studies teachers (the National Council for the Social Studies or NCSS). I intended to survey both history and social studies teachers in the state of Massachusetts and did not want to limit my sample to members of a professional organization. For that reason, our sampling procedures differed tremendously. I address sampling procedures in greater depth below.

To establish reliability of the survey instrument, Vinson did a repeat mailing pilot study. The researcher sent the initial survey to a convenience sample of sixty five high school teachers in large, suburban school districts. Thirty teachers responded to the first administration of the survey. One month later, Vinson repeated the mailing to the thirty teachers. This time, sixteen teachers responded to the survey. The test-retest pilot study then was based upon the first and second administrations of the survey to a convenience sample of sixteen. Vinson obtained reliability coefficients for both calculated and selected instructional approaches between the two administrations of the instrument to the sixteen participants. Here, he defined the reliability coefficient as "the proportion of participants' assigned identical classification [of Calculated Orientation] in each application of the questionnaire" (p. 122). He found that for Calculated Orientation, " $r_{xx} = .50$ (i.e. eight of sixteen respondents received the same calculated instructional approach category in each pilot administration)" (p. 122 - 123). Vinson relied on additional indicators to assess reliability of his survey, including responses to a survey asking

participants in the two pilot studies to identify confusing terms. He also assessed reliability based on rates of non-response to individual survey items. Vinson's reliability study led the author to the conclusion that the survey instrument was "at least moderately reliable" (Vinson, 1996, 127). Aside from the dearth of data collected for these reliability studies and the relatively unconvincing results of Vinson's reliability analyses, his approach to establishing reliability is nonconventional. For this dissertation, I plan to conduct scale reliability analyses, factor analyses, and other statistical procedures to establish the reliability both of the scales and of the instrument as a whole.

First Revisions Made to the Survey Instrument

Vinson's instrument was not enough to answer some of the questions of this study. While his instrument was able to get at teachers' approaches to history and social studies, it was not adequate to capture teachers' reactions to the 10th/11th grade U.S. History MCAS test. For that reason, I selected a few items from a second instrument to examine teachers' reactions to state-mandated assessments. Pedulla et al's (2003) empirical study (discussed in the literature review) surveyed teachers' opinions and beliefs about state-mandated assessments and provided evidence about teachers' beliefs. Questions for my revised survey instrument included items from both the Vinson study as well as Pedulla's (2003) study. Specifically, I used their items classifying teachers' background information including years of teaching experience, gender, age range and race. I also included one item on the survey that asked teachers about their political orientation (see Appendix C). Additionally, I used six items from the Pedulla et al. study that asked about teachers' beliefs about standards and standardized testing.

I wanted to use previously administered surveys in order to avoid some of the measurement errors and limit problems with reliability, validity and other risks associated with construction of my own instrument. However, I needed to make a series of structural changes to the survey instrument to offset some of the problems Vinson encountered. An examination of Vinson's instrument shows that the researcher may have unwittingly allowed some instrument error in the administration of his survey. First, in Vinson's instrument, all of the questions meant to calculate one teaching approach are at the beginning of the survey and all of the items meant to calculate another teaching approach are toward the end of the survey. The researcher did not appear to take into account the possible effects of survey fatigue in that people may have rated less favorable the approaches found toward the end of the survey. The results of Vinson's survey show that teachers rated less favorably the content, pedagogical and philosophical choices associated with the cultural transmission teaching approach. Participants favoring the cultural transmission approach would have found four out of six items meant to describe their preferred approach on the very last page of the survey. To correct this oversight, I revised the instrument so that the items that align with each teaching orientation are randomly distributed throughout the survey. By distributing a revised version of the instrument with randomized items, I was hoping to avoid some of the instrument error that Vinson introduced in his administration of the same survey.

Table 1: *Breakdown of the Instrument by Item Type*

	Cultural Transmission	Social Sciences	Reflective Inquiry	Reflective Inquiry for Social Ends	Reflective Inquiry for Personal Ends
Purpose	Item # 09 Item # 16	Item # 10 Item # 13	Item # 07 Item # 15	Item # 08 Item # 11	Item # 12 Item # 14
Content	Item # 19 Item # 26	Item # 21 Item # 23	Item # 18 Item # 20	Item # 17 Item # 22	Item # 24 Item # 25
Methods	Item # 28 Item # 36	Item # 30 Item # 32	Item # 31 Item # 33	Item # 27 Item # 34	Item # 29 Item # 35

In the Vinson instrument, after each item, participants were provided with numbers one through five and were asked to circle the one numbered response “that most closely represents [the participant’s] belief with respect to each item.” On recommendation of Dillman (2000), I amended Vinson’s Likert scale so that it only included four answer options and I labeled each of the choices for each (Strongly Agree, Agree, Disagree, Strongly Disagree) so that there would be no ambiguity for the participant about the choice they were making. I amended the directions so that they were clearer. I changed the formatting of the survey so that all items to do with the

participants' perceptions about the purpose of HSS were in one section; their perceptions about pedagogical choices were in the next section; and their perceptions about appropriate content choices were in the next section. Separating the items into sections gave increased visual clarity to the survey instrument. Finally, I made a number of edits to Vinson's items – particularly items which had low or no correlation in his administration of the instrument. Specifically, I removed extraneous phrases, tried to limit double barreled questions and assured that items that were intended to be highly correlated were actually theoretically in line with one another. For example, in Vinson's instrument there were two items that were intended to get at teachers who were reflective inquirers. Reflective inquirers, in theory, align most of their instruction to the needs and desires of their pupils. However, in this instrument one of the two items that was meant to be highly favored by these teachers had the stem, "teachers should." For items such as these, I took out references that would lead the participant to believe that the teacher was at the center of instruction rather than the student.

Pilot Studies and Second Revision Made to Survey Instrument

After these changes were made to the survey instrument, two other techniques were used to increase the reliability and validity of the survey instrument. First, I conducted a read aloud with a volunteer who previously taught history and social studies for six years. The volunteer read questions aloud, responded to questions aloud and then made answer choices aloud. The read aloud volunteer's feedback was used to amend the survey instrument a second time. Some key changes included creating greater clarity in the directions, and the insertion of examples of "social science disciplines" in related

items (i.e. “historians, geographers, political scientists”). I asked the read aloud volunteer what distinguished answers for her between “strongly agree” and “agree” or “strongly disagree” and “disagree.” She noted that when she agreed with an item, she likely would endorse another teacher who used that approach in the classroom, but that she didn’t necessarily use that approach. However, when she “strongly agreed” with an item, it usually meant that she tried actively to do those things in her own classroom. The reverse was true with the disagree items. When she “disagreed” with something, she would not endorse a teacher doing those things in their classroom. When she “strongly disagreed” with something, she actively tried to avoid those practices or actively discouraged colleagues from doing those things in their classrooms.

Next, like Vinson, I piloted the survey instrument with a small sample of convenience of social science instructors to isolate and correct additional problems with the instrument. I asked the pilot group to note confusing wording, double barreled questions, and problems with item stems or other obvious problems on the survey. One recommendation that these pilot study participants made (in addition to the read aloud volunteer) were comments about their desire for a middle, “do not know,” “not applicable” or “neutral” option. A second set of comments was about whether participants should use a “check” or an “x” to mark their answer. Again, this comment was also made by the read aloud volunteer. The instrument was altered to read, “Please indicate the strength of your agreement or disagreement with each of the following statements by marking the corresponding box with an ‘x.’”

A final set of comments was made by the pilot group about items that were double barreled, or in the case of items 37 through 41, had a lot of information to process at once. I reduced the number of items that were double barreled as much as possible. With a few items, however (e.g. “High School social studies content should emphasize the importance of minority groups”) key terms like “emphasize” that made the item double barreled were left in. This was because, theoretically, the item was designed to be strongly favored by a particular orientation in light of the inclusion of those specific terms. Based on these comments and recommendations, I revised the instrument a second time and prepared and submitted an application to Boston College’s Institutional Review Board (IRB). The study was approved by the IRB on February 4th, 2009.

Research Sample

Though many surveys of the HSS field use membership lists in the National Council for the Social Studies as a population frame (Barr, et al., 1977; Vinson, 1998), other scholars have argued that data gathered from this population frame are considerably skewed to favor more liberal and critical teaching orientations (Leming, 1992). As mentioned previously, Vinson’s study was administered to a sample drawn from this population frame.

For my study, I used all public secondary high schools that instruct 10th and 11th grade students in history or social studies in the state of Massachusetts (N = 352) as a population frame. I obtained school demographic data that was current for the 2008-2009 academic year from the Massachusetts Department of Elementary and Secondary Education (MassDESE) website. From this frame, I constructed a random sample of

schools. Data files for history and for social studies teachers for each of the 352 schools were merged into one file and then sorted based on the size of the HSS department within the school. Schools were then stratified into four quartiles based on the size of the HSS department. Stratifying the sample this way helped me target both large, urban and suburban districts and smaller, rural districts. Random numbers were generated for each of the schools and quartiles were then sorted based on their random number.

The number of schools sampled and the number of teachers that needed to be sampled were determined by power analyses based on the population size of schools ($N=354$) and history and social studies teachers in those schools ($N=2,328$). Based on an a priori computation for required sample size conducted with G*Power software, in order to conduct t-tests of difference between two independent means with power set at .95 and population error probability set at 0.05, my sample size for each group should be 105 participants. Assuming five orientations exist, I needed a sample size of approximately 525 teachers for appropriate power. All other proposed statistical tests (frequencies and Chi-Squared procedures) required smaller sample sizes than the t-test of independent means. Since 2,328 teachers were found at 354 schools, I deduced that each school would have an average of seven teachers in the HSS department. This meant that my sample should include approximately 19 schools from each quartile with an average of seven teachers ($n=133$) from each quartile to get approximately 525 teachers. Teachers from small history and social studies departments would have been underrepresented using this sampling technique, however, I reasoned, more teachers work at larger schools. While it would be ideal to use all secondary HSS teachers in the state of Massachusetts as my

population frame, the MassDESE does not make name and contact information for each teacher available. It was therefore the next best thing to work with the schools as the population frame. Using schools as a population frame to sample schools and then teachers has been done in at least one previous, high profile survey (NEA study in Leming, 1991).

Using the population frame of the 352 schools that were sorted by department size and then randomly ranked, I selected the first twenty schools in each of the four quartiles (n=80) for the study. Once schools from each quartile were randomly selected through this process, I visited each of the 80 schools' websites or called each of the schools to collect contact information for all of the members of the history and social studies department. When names were not available, I made contact with department chairs to try to obtain more detailed information. When this process was completed, I had a sample frame that included 665 teachers from 80 different schools. After searching fruitlessly at two schools for the names of department members or even an indication that the school had teachers who taught the subject, I decided to remove two schools from the sample.

The number of participants in the study was slightly higher than the power analysis called for. This was due in large part to the fact that it was difficult at some schools to weed out the teachers who did not teach U.S. History as one of their subjects. In other words, some schools' websites made it clear the names of teachers who taught U.S. History and others just had a list of department members. This meant that my final sample included some teachers who did not teach U.S. History.

Survey Administration

The survey was revised and administered using the recommendations and guidance of Dillman's (2000) *Tailored Design Method*. In terms of survey administration, the author made many recommendations for effectively administering surveys to minimize both item non-response and unit non-response error – thereby increasing the validity of the survey. A postcard indicating that the survey was forthcoming was mailed to all department chairs and all faculty in schools that had no department chairs in February 2009. Two weeks later, packets were mailed to department chairs, or directly to faculty in schools with no department chairs. These packets contained a cover letter explaining the intent of the survey, protections those participants were afforded as well as a link to the electronic version of the survey (see Appendix D). The packets also contained envelopes for each teacher in the department. The envelopes contained cover letters addressed to the participant, a paper copy of the survey with coded "Random IDs" to track participation, and a self-addressed stamped envelope that participants could use to return the survey. Packets were mailed to department chairs in an effort to boost response rate and credibility of the survey. Meanwhile, individual envelopes addressed to each faculty member were placed in the larger envelope to ensure participant's confidentiality. About four weeks after the initial questionnaire was mailed to participants, a postcard was mailed directly to participants who did not respond. On this postcard, I directed participants to the online version of the survey. Two weeks after the second notification was sent, I made one final email contact with the participants.

In addition to tailoring the survey administration, there were a number of recommendations Dillman (2000) proposed for boosting response rates to the survey. First, participants were told from the beginning that, if they wish, they would be able to see the results of the survey. They were instructed to check a box indicating their interest and asked to return the survey within two weeks. An email was sent to respondents who indicated their interest in seeing results of the survey that directed them to an executive summary of the dissertation findings on the researcher's homepage.^x Making survey results available to survey participants created an incentive for participants to complete the survey. I also included in the survey a token of appreciation – participants were informed that once their completed surveys were returned they were entered into a drawing for three gift cards to *Barnes and Noble* bookstore. Both of these rewards (allowing access to findings and including tokens of appreciation) were designed to invoke in participants the norm of social exchange and to increase the likelihood of response.

Using this sampling technique, 665 surveys were mailed out. Fifteen people responded noting that they did not teach anymore or they did not teach U.S. history or at the high school level. Removing these people from the total sample left 650 people. 272 people returned their surveys for a response rate of 42%.

Data Analysis

To be sure that I used the correct techniques to analyze the data, I returned to the original question that the survey was meant to address. I created a table with the main question and several sub-questions that needed to be answered with survey results. The

survey was meant to answer the question, “What are the orientations of history and social studies teachers in the state of Massachusetts?” This question implied a few sub-questions. These sub-questions pointed to specific statistical analyses I needed to conduct to answer the sub-questions. In Table 2, I outline the sub-questions and the statistical analyses I conducted to answer the questions.

Table 2: *Proposed Research Questions and Analysis Techniques*

Sub-Question	Analysis Technique
What are the demographic characteristics of the sample?	Frequency analysis
Do sample demographic characteristics correlate with national population?	Compare to known demographics with Chi-Squared analysis
Do the data demonstrate distinguishable orientations for teaching history and social studies?	Factor analysis and scale reliability analysis
Do the teaching orientations adhere to current theory that posits five teaching orientations?	Scale reliability analysis
Do survey items intended to measure each of the five orientations hang together in a predictable manner that matches theory?	Factor analysis

Do the six items that represent “calculated orientation” correlate with the one item that represents “selected orientation?”	Chi-Squared analysis
What are the individual teachers’ mean scores for each of the instructional approaches?	Calculated mean scores for each orientation
Is there one mean score that is statistically significantly higher than the other mean scores such that one orientation best represents that teacher’s approach?	Repeated measures ANOVA Statistical ties analyzed for patterns
What are the frequencies of calculated orientations in the state of Massachusetts	Frequency analysis
Do specific teaching orientations correlate to the favorability rating of the 10 th and 11 th grade state mandated assessment in history and social studies?	Chi-Squared analysis
Do specific teaching approaches correlate to particular political ideologies?	Chi-Squared analysis

Data were loaded into SPSS software. The data were cleaned and coded. Initial descriptive statistics were calculated to analyze the demographic characteristics of those who responded to the survey to establish whether or not the sample was representative of HSS teachers by noting whether or not the sample is similar to or differs from previous

studies of HSS teacher characteristics (Bliss & Banks, 1994; Leming, 1991; Ochoa, 1981). Though no studies exist that describe the history and social studies teaching force of Massachusetts specifically, general comparisons were made to the national population in terms of race, gender, professional training and years in practice to the findings of a previous study (Leming, 1991). Although Leming's review was based on surveys from as far back as 20 years ago, it remains the most current and comprehensive review to date of the demographic characteristics of HSS teachers.

Next, inferential statistical techniques were conducted including factor analyses, to see if the items loaded similarly to studies for which the original instrument was used. Results from this study were compared to White's (1982) validation study of the Barr, Barth and Shermis survey instrument. The data did bear out the theoretical assumptions of earlier researchers.

Individuals' mean scores for each teaching orientation were calculated to determine if the teachers' mean scores for each orientation were statistically significantly different from their mean scores for the other teaching orientations. Similar to Vinson, I assigned teachers to categories of instructional approach based upon their highest mean score. Teachers whose scores represented a tie were categorized as "other." Once the teacher's instructional approach category was assigned, I did a frequency analysis to determine the proportion of teachers in each of the teaching orientations. For example, did those whose answers reflected a "critical thinking" orientation comprise the majority of HSS teachers in the state of Massachusetts as was found in Anderson et al's (1997) Q-Sort and Vinson's (1998) study? Or, did the more conservative cultural transmission

approach have a better representation in this study than in previous studies as Leming (2006) contended?

Once I established these characteristics of HSS teachers, I needed to explore whether or not there were any statistically significant relationships between HSS teachers' orientations and their reaction to Massachusetts standards and assessments on history and social studies. Statistically this was represented by a significant correlation between the orientation of social studies teachers and items on the instrument that asked about their beliefs about the U.S. History MCAS test (i.e. modified items from the Pedulla et al study). Vinson used multiple statistical analyses to find correlations between a teacher's instructional approach and their power/knowledge categories including a bivariate crosstabulation. I performed a similar statistical procedure however, in lieu of "power/knowledge" categories, I looked at teachers' responses to items that asked about the MCAS U.S. History test.

I also attempted to replicate or better the validation findings of Vinson's original study. I conducted my own validation studies including scale analyses and factor analyses to ascertain how well the six items used to measure each instructional approach worked together. Additionally, I conducted the same simultaneous cross-tabulation procedure to analyze the relationship between calculated and selected instructional approach.

Design II: Massachusetts History and Social Studies Standards and Assessments

Data Collection

To explore the aims of the state standards and assessments in history and social studies, I focused specifically on three different documents, all of which can be found on

the MassDESE website. First, I looked at the Massachusetts Department of Education's MCAS Guide to History and Social Studies Assessments (*MCAS Guide to History and Social Science Assessments*, 2007). This document includes a description of the development of the high school U.S. history assessment as well as a sample test of the high school (usually occurring in grade 10 or 11) standardized United States History MCAS (Massachusetts Comprehensive Assessment System) test.

The sample test included in the MCAS Guide to History and Social Studies Assessments is the second document that was analyzed. Though the sample test was not an actual, full-length version of the United States history assessment, the MCAS guide argued that the sample test items were representative of the actual test that high school students were going to take in spring 2009. The sample test included examples of each of the different types of test items as well as a large section on multiple choice items, and smaller sections with items that utilized time lines, maps, graphs, and long documents such as the *Bill of Rights* and the *Gettysburg Address*.

The third document that was analyzed was another collection of sample test items provided by the Massachusetts Department of Elementary and Secondary Education (*High School U.S. History Released Items*, 2007). The two sets of sample test items were compiled together (Appendix F) and were the centerpiece of the content analysis described in chapter five. All three of these documents (the MCAS Guide to History and Social Studies Assessments and the two sample tests of high school U.S. history) can be found on the Massachusetts Department of Elementary and Secondary Education website

(www.doe.mass.edu). All three documents were analyzed as discussed in the next section on data analysis.

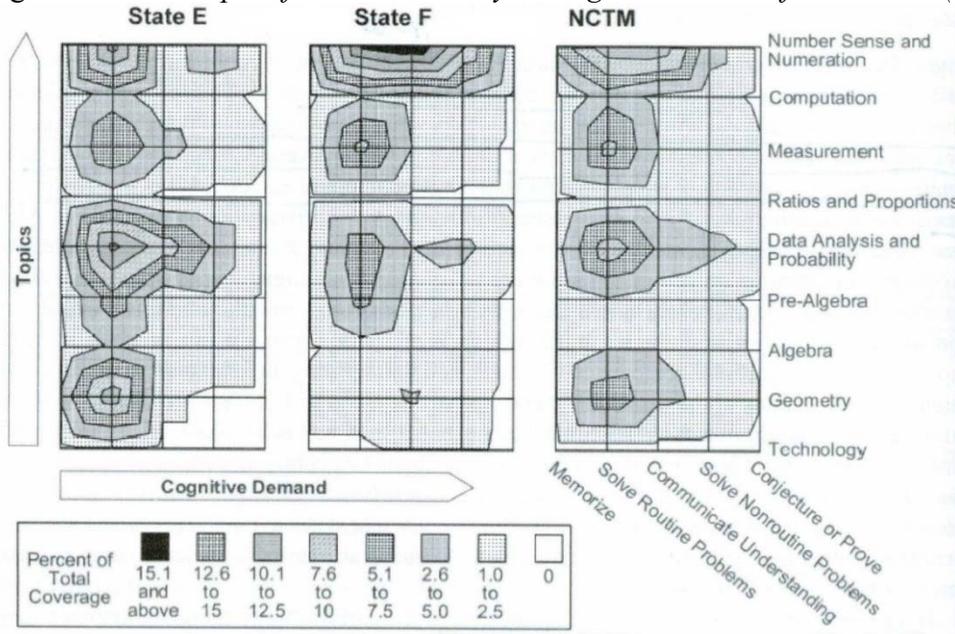
Data Analysis (Content Analysis)

Bazerman (2006) wrote that “written texts pervade the educational process, the educational system, and the policy and practical processes that shape education” and education researchers have various reasons for analyzing texts (p. 77). Systematic analysis of educational texts, therefore, is addressed throughout education research literature. For this study, I analyzed the MCAS assessments in history and social studies to understand better the practical consequences of how the assessments linked to teachers’ orientations (p. 77). Researchers select from multiple methods of text analysis depending on which method best allows them to analyze their particular research problem. Bazerman noted that content analysis is a specific form of text analysis that helps researchers systematically examine what appears in the text itself in order to interpret the meaning of the text. In this method, the researcher first identifies kernels of words, terms or categories of content and then examines the text(s) in search of occurrences of that content (p. 83). One downfall of content analysis is that it “treats the meaning as unproblematic and directly revealed through the words” (p. 83). Despite this valid critique of the method, it was best aligned to the purposes of this study in the sense that it enabled the raters to analyze dozens of pages of text within a manageable time frame.

As policy actors and the public increasingly call for state-mandated tests with higher and higher stakes attached, systematic analyses of assessments and their alignment

with what is taught in the classroom becomes critical. Porter (2006) outlined a particular content analysis method to assess alignment between the enacted curriculum (taught), and the assessed curriculum (tested). First, a researcher must define the “content language.” Content language can include all of a subject’s concepts outlined as specific knowledge and skills in states’ curriculum standards. However, Porter contended that doing content analyses based on curricular content is “insufficient in explaining variance in student achievement” (p. 142). Curricular content must be analyzed alongside the concept’s cognitive demand – a measure oftentimes referred to as level of thinking or depth of knowledge. To analyze the occurrence of, and the relationship between, content language and cognitive demand, Porter explained that researchers can develop a matrix with the cognitive demand on the x axis and content language (content topics) on the y axis (see Figure 1). For the content analysis component of this study in which I analyzed the Massachusetts assessment for 10th and 11th grade U.S. history, I followed this model of content analysis. I created a similar matrix to analyze the occurrence of content and cognitive demand of the Massachusetts curriculum assessment for 10th and 11th grade history and social studies.

Figure 1: An Example of a Content Analysis Alignment Matrix from Porter (2006)



As mentioned previously, an analysis of the standardized test entailed an examination of how specific test items tapped varying levels of cognitive demand. Yet, there were several typologies of cognitive demand to choose from. Kevin Dwyer, Lead Developer for the U.S. history MCAS test at the Massachusetts Department of Education revealed that state-level analyses of the U.S. history assessment were done using Norm Webb’s taxonomy of Depth of Knowledge (DOK). This taxonomy has four levels: (a) Level 1: Recollection; (b) Level II: Basic Reasoning; (c) Level III: Complex Reasoning; and (d) Level IV: Extended Reasoning. Each level, and each level’s relevance to history and social studies subjects in particular, is described in greater detail throughout Webb’s scholarship. Based on Webb’s work, these four levels of depth of knowledge were plotted along the x-axis of my content analysis matrix.

This content analysis required a framework for the content categories along the y-axis of the matrix. To conduct a content analysis that would help answer the research questions posed in this study, I chose a framework for the content language that was in line with theories in HSS about dominant curricular orientations. The content descriptions for each of the orientations are included in the review of the literature on teachers' orientations in HSS. For example, a CT teacher may choose to transmit content that reflects traditional American values and symbols. A SS teacher may choose content that teaches students particular skills such as data gathering techniques. The five orientations and the content that teachers within each of the five orientations theoretically emphasize were plotted along the y-axis of the matrix.

A panel of HSS experts was composed and was given the task of reviewing each item on the sample MCAS-US test and assigning the item to one DOK category and one content category. Once the items of the test were compiled in this matrix format, two separate techniques were recommended for determining the alignment between the assessment and teacher's orientations in the state of Massachusetts. First, Porter (2006) outlined a method for mathematically determining alignment based on a calculation of proportionality, cell by cell, of the standards versus the assessment and, in this case versus the occurrence of teachers' orientations. This is described in greater detail in chapter five.

Porter's method (Porter & Smithson, 2001) served as a starting point in the analysis of the relationship between HSS teachers' orientations in the state of Massachusetts and the state's assessment related to HSS. Simply put, the null hypothesis

was that there was no significant difference between the proportions of items on the assessment meant to measure content favored by different orientations to history and social studies and the proportion of teachers claiming to teach from that orientation. The alternative hypothesis was that there is a significant difference in proportions. For example, teachers who claimed to teach from the cultural transmission orientation might only comprise 15% of the sampled teachers in Massachusetts but the content analysis of the test might show that 70% of the MCAS-US sample items test content favored by the CT orientation. Porter's techniques for content analysis aided in the comparison of proportions. I refer to Porter's technique as a starting point because I intended to analyze and discuss the nature of the relationship between the aims of HSS teachers and state standards and assessments related to history and social studies in a more holistic manner.

Information gathered in preparation for this study has already been shared with standards and test makers at the state level in Massachusetts. Findings from this study may serve to inform education policy in the state of Massachusetts including establishing or contradicting both the value and the validity of this particular assessment instrument.

In summary, this dissertation was designed to inform the debate about the continued poor performance of history and social studies students on standardized history tests and rates of civic participation through a systematic study of the orientations of history and social studies teachers in Massachusetts, the skills and constructs measured by the MCAS-US history test, and the relationship between the two. This study examines this complex relationship via two research designs. First, a survey of Massachusetts history and social studies teachers was conducted to analyze the orientations from which

teachers approach the subject. Second, a content analysis of the MCAS-US test was conducted to identify the skills and constructs assessed on the test. Both the survey and the content analysis were carried out through the theoretical lens of democratic pragmatism, and both employed the same framework for understanding the varied ways that history and social studies is taught. This allowed for both quantitative and qualitative analyses to be done on the third research question posed in this dissertation. In chapter four and chapter five, the findings of these two research designs are presented.

CHAPTER FOUR: THE NUANCES OF MASSACHUSETTS HSS TEACHERS

In the first chapter of this dissertation, I discussed the problem of declining civic participation in the United States, noting that many believe that public schools can curb this decline by initiating programs to boost students' performance in history and social studies. Critics of these educational efforts claim that they are failing to turn the tide of civic disengagement, pointing to poor student performances on standardized history tests as evidence and blaming teachers for failing to teach students meaningful content or important history. Others, however, counter that students' poor performance on these tests signal a problem with the test's construction and content rather than with teachers. They question whether test makers have lost sight of the true purpose of history and social studies and call for alternative measures of civic activism and achievement.

The next three chapters present analyses of the data collected for this study. This chapter is an analysis of a systematic survey of Massachusetts history and social studies teachers, which addresses the question, "What are the curricular orientations of history and social studies teachers in the state of Massachusetts?" In this chapter, I make two key arguments. First, I argue that in order to understand HSS curricular orientations, a more nuanced interpretation of teachers' stances within those orientations, which encompasses but goes beyond the three traditions thesis, must be explicated. Second, I argue that teachers in Massachusetts espouse a relatively uniform message about the purpose, pedagogy and content of HSS – one that is noticeably in line with democratic pragmatism.

Analysis of the Survey Instrument: Five or Three Orientations of HSS Teachers?

For decades, educational researchers have been working to understand and classify the various orientations of history and social studies (HSS) teachers. To briefly recap my discussion in chapter two, in the late 1970s, Barr, Barth and Shermis theorized that there were three distinct and reliable traditions that most HSS teachers align with: (a) the cultural transmission orientation (CT), b) the social studies as a social science orientation (SS) and, c) the reflective inquiry orientation (RI).

A critical review of the current literature suggested that the RI tradition was more nuanced and contained more discrete categories than Barr, Barth and Shermis originally postulated. Specifically, there were several studies that created a separate category for HSS teachers who used inquiry methods for the purpose of promoting social justice and equity. Additionally, there were several studies that created another separate category of HSS teachers who used inquiry methods to help their students achieve personal ends (i.e. values clarification and personal growth). For both of the categories, inquiry was the preferred pedagogical approach. What differentiated the categories was that inquiry was used as a means to achieve different ends (social or personal). Therefore, I broke the RI orientation into three categories: a) reflective inquiry (RI), b) reflective inquiry for social ends (RISE), and c) reflective inquiry for personal ends (RIPE). The decision to break the RI orientation into these three separate orientations created a total of five distinct categories of teaching orientations and was a break with previous empirical studies. In my study, teachers' orientations were expected to reflect five orientations. The five orientations included Barr, Barth and Shermis' original three (CT, SS and RI), with two

additional categories describing specific goals of the teachers working from an RI orientation (RIPE and RISE). The survey instrument used in this study (a modification of Vinson's 1996 survey instrument) was based on the framework of five, rather than three distinct orientations. To determine the instrument's validity and reliability, and to ascertain whether the modifications I made actually improved the instrument, a number of statistical tests were performed.

I found Massachusetts teachers largely within Barr, Barth and Shermis' original three orientations (CT, SS, RI). In other words, while I believed that the theory of three orientations could be improved upon by further dividing the three orientations into five, the statistical analyses of the survey data showed that the original three orientations worked best for this sample. I supplement this finding however, by arguing for an alternative understanding of orientations – one that reflects gradations, nuances and overlap between and among orientations that teachers animate in the classroom.

Scale Reliability Analysis

Scale reliability analyses were conducted on the scales intended to measure each of the five orientations (CT, SS, RI, RIPE and RISE). Each scale had a total of six items, with three subscales intended to measure what participants believed to be the appropriate purpose, method and content for each of the five orientations. Scales that were intended to measure each of the five orientations yielded Cronbach's α ranging from .607 (moderate) to .776 (moderately strong) (see Table 3).

Table 3: *Scale Reliability Analysis of Items Measuring Five Orientations*

Curricular Orientation	Cronbach's Alpha (α)	Mean Inter-Item Corr.
Cultural Transmission (Total)	.607	.214
Purpose	.290	.191
Methods/Pedagogy	.509	.342
Content	.297	.178
Social Science (Total)	.672	.256
Purpose	.537	.368
Methods/Pedagogy	.486	.321
Content	.727	.573
Reflective Inquiry (Total)	.681	.260
Purpose	.547	.377
Methods/Pedagogy	.505	.339
Content	.810	.681
RI for Social Ends (Total)	.776	.363
Purpose	.453	.297
Methods/Pedagogy	.519	.359
Content	.774	.632
RI for Personal Ends (Total)	.764	.343
Purpose	.676	.527
Methods/Pedagogy	.555	.390
Content	.616	.616

Definitions of strength of scale reliability scores vary, but generally statisticians prefer to see a Cronbach's α of at least .7. Using this standard, the scales were not highly reliable but do show some ability to measure a given concept. Interestingly, the scale with the lowest reliability in this study was the scale intended to measure the cultural transmission orientation – a scale that is usually found to be the most reliable in previous studies. I address this anomaly in greater depth below.

It should be noted that when I collapsed the items that were intended to measure RI, RISE and RIPE orientations into one single scale, Cronbach's α jumped from $\alpha = .681$, $.776$ and $.764$ respectively to $\alpha = .886$ (Table 4). This might suggest that Barr, Barth and Shermis' three traditions model captures this sample of HSS teachers more reliably than the proposed five traditions alternative. However, because Cronbach's α scores of scale reliability tend to increase when items are added to a scale, it is highly likely that the increase in Cronbach's α scores was due to the number of items on the scale increasing from six to eighteen. Thus, more evidence is needed to determine that the three traditions model more accurately captures the orientations of Massachusetts HSS teachers than the five traditions alternative I proposed.

Table 4: *Scale Reliability Analysis of Items Measuring Three Orientations*

Curricular Orientation	Cronbach's Alpha (α)	Mean Inter-Item Corr.
Reflective Inquiry (Including RI, RIPE, RISE)	.886	.229
Purpose	.763	.353
Methods/Pedagogy	.708	.294
Content	.839	.468

Meanwhile, reliability measures of subscales composed of the two items in each orientation intended to measure content, methods and purpose, yielded Cronbach's α ranging from as low as .29 to a high of .81 (see Table 4). With only two items in the subscale, it is difficult to achieve high reliability. However, because previous studies have done reliability analyses on these subscales, I attempted to replicate these other studies. Pallant (2009) suggested that when conducting reliability analyses on scales of six items or less, researchers should also examine mean inter-item correlations to determine scale reliability. She argues that mean inter-item correlations should fall between .2 and .4. Using Pallant's standard for acceptable mean inter-item correlations together with the standard for acceptable Cronbach's α , each of the scales and subscales in this instrument show moderate to strong reliability with only a few notable exceptions.

One exception is with the subscales intended to measure various aspects within the cultural transmission orientation. The subscales – or two items within each scale – intended to measure the “purpose” of the subject, and the preferred “content” of HSS

teachers within the CT orientation are not reliable by either the Pallant standard or the Cronbach α standard. This is a surprising finding given that, according to the literature, the CT orientation is the most consistent and easily measured orientation nationally. One possibility is that the four items which comprise the two subscales do not accurately reflect the beliefs of teachers working from a CT orientation, or they were worded awkwardly. However, after examining the items in these subscales once again for word choice and clarity, it is apparent that these items are clearly worded and straight forward. These items are also exact replications, or only slight modifications of items used to measure the CT orientation on previously administered surveys dating back to Barr, Barth and Shermis' original implementation in 1977. In other words, the low scale reliability of these items is not attributable to the item itself. The use of the same or similarly worded items in a survey to a nationwide sample has proven reliable in the past.

A second possibility as to why this scale had low Cronbach's α is that some of the items on the CT scale are items that might be favorable to most HSS teachers (e.g. "As a result of high school social studies, students should be able to demonstrate knowledge of traditional American history, geography, culture, and values."). In a case where most teachers in the sample responded favorably to one or two items on the scale, the scale may not differentiate teachers from one another enough to produce high Cronbach's α . This is a likely deduction, particularly in light of the fact that the CT orientation is usually the most easily defined and readily observed tradition of HSS. In other words, when similar items were used in previous instruments, more teachers responded positively to these CT items in nationwide studies than in this study of Massachusetts

teachers. The low reliability of these subscales is most likely a reflection of the peculiarities of this sample of Massachusetts teachers rather than either a poor subscale, or a nationwide decline in CT oriented HSS teachers. This is discussed further in following sections where I analyze the data and compare the Massachusetts HSS teachers in this sample to the nationwide population.

Two other subscales that had low reliability based on Cronbach's α and the mean inter-item correlation standard were the subscales intended to measure the "purpose" and the "content" of the proposed RIPE orientation. This finding in combination with the t-test and repeated measures ANOVA findings that there is not a statistically significant difference between individual's RI and RIPE scores (addressed below) provides further evidence that Barr, Barth and Shermis' three traditions are more reliably measurable than the proposed alternative five traditions. Once the RIPE items were collapsed into the larger RI category, reliability of these two subscales increased from $\alpha = .676$ and $.616$ to $\alpha = .763$ and $.839$ respectively.

One final check of internal validity was done to see if scales were reliably measuring teachers' orientations in HSS. The five scales representing the five theorized orientations (CT, SS, RI, RISE, RIPE) were comprised of six items each for a total of thirty questions. As stated previously, teachers were assigned a "Calculated Orientation" based on their highest score across those five scales. To assess whether their calculated orientation reliably represented teachers' orientations, respondents were asked to read through a summary of each of the five orientations and chose the one with which they most agreed. This became the teachers' "Selected Orientation." I conducted a Chi-Square

test of association with an $\alpha = 0.05$ to determine whether there was a statistically significant relationship between the two variables “Calculated Orientation” and “Selected Orientation.”

When the test was run assuming the five orientations, several of the cells had expected counts fewer than five (>60%) thus artificially inflating the significance of the Chi-Square statistic. The number of orientations was then collapsed down to three orientations (the three categories measuring reflective inquiry were collapsed into one category). The number of cells with expected counts less than five still remained too high for accurate analysis (>44%). The orientations were then collapsed a final time based on White’s (1982) contention that there are in fact only two orientations in HSS – a conservative “CT/SS” orientation and a liberal “RI” orientation. When both calculated and selected orientations were collapsed to reflect the theory of two orientations, expected counts were greater than five in each of the cells providing an opportunity for accurate interpretation.

Results of the Chi-Square analysis (Table 5) indicate that there is a statistically significant relationship between the calculated orientation and selected orientation ($p < .000$). The null hypothesis that there is no relationship between teachers’ selected orientation and teachers’ calculated orientation was rejected. Specifically, participants whose “calculated orientation” was either CT or SS were statistically significantly more likely to choose CT or SS as their “selected orientation” and statistically less likely to choose RI as their “selected orientation.” Based on this information and the scale

reliability analysis above, it appears as though the 30 items intended to measure calculated orientation accurately reflected teachers' orientations in HSS.

Table 5: *Internal Validity Check of Calculated Versus Selected Orientations Using Chi-Squared Analysis*

Selected orientation	Calculated Orientation		X ²
	RI	SS/CT	
RI	155 (1.2)	10 (-2.9)	34.940*
SS/CT	39 (-1.9)	23 (4.7)	

Note. * = $p < .05$. Adjusted standardized residuals appear in the parentheses below group frequencies.

Repeated Measures ANOVA Test

Next, repeated measures ANOVA tests were conducted to determine whether individual's scores for each of the five orientations were statistically significantly different from one another so as to suggest that individuals had a strong preference for one orientation over another. Data from this test is presented in Table 6 and Table 7.

$$H_0 : \mu_{CT} = \mu_{SS} = \mu_{RI} = \mu_{RIPE} = \mu_{RISE}$$

$$H_A : \mu_i \neq \mu_j ; \text{ for some } i+j = CT, SS, RI, RIPE, RISE$$

Using the Huynh-Feldt correction, $\varepsilon = .742$.

Table 6: *Repeated Measures ANOVA Test on Scores of Five Orientations*

	Type III	df	Mean	F	Sig.
	Sum of		Square		
	Squares				
Huynh-Feldt	52.034	2.967	17.538	106.255	.000

Table 7: Repeated Measures ANOVA Test: Relationships between Orientations

Total Calculated Scores	Total Calculated Scores	Mean Difference	Std. Error	Sig.
CT	SS	-.252*	.034	.000
	RI	-.394*	.037	.000
	RISE	-.618*	.039	.000
	RIPE	-.356*	.038	.000
SS	RI	-.141*	.026	.000
	RISE	-.366*	.029	.000
	RIPE	-.104*	.031	.009
RI	RISE	-.225*	.023	.000
RISE	RIPE	.262*	.024	.000

I rejected the null hypothesis at the .05 level. I found statistically significant differences between the means of all but the same pair of total calculated scores. Specifically, the CT mean score (2.676) was significantly lower than all of the other scores. While higher than the CT mean score, the SS mean score (2.929) was lower than each of the three RI family mean scores. The RI mean score (3.070) and the RIPE mean score (3.032) were the only pair of mean scores that were not statistically significantly different from one another. Finally, the RISE orientation's mean score (3.294) was the highest of the five scores. The

finding that the mean total scores were each statistically significantly different from one another except the RI and RIPE orientation demonstrates further support for collapsing the five orientations into three.

I next conducted a repeated measures ANOVA test with only three orientations including the collapsed RI orientation to determine whether teachers' scores for each of the three orientations were statistically significantly different from one another so as to suggest that individuals had a strong preference for one orientation over another. Data from this test is presented in Table 8 and Table 9.

$$H_0 : \mu_{CT} = \mu_{SS} = \mu_{RI}$$

$$H_A : \mu_i \neq \mu_j ; \text{ for some } I+j = CT, SS, RI, RIPE, RISE$$

Using the Huynh-Feldt correction, $\varepsilon = .888$.

Table 8: *Repeated Measures ANOVA Test on Scores of Three Orientations*

	Type III	df	Mean	F	Sig.
	Sum of		Square		
	Squares				
Huynh-Feldt	26.799	1.775	15.096	102.201	.000

Table 9: *Repeated Measures ANOVA Test: Relationships between Three Orientations*

Total	Total			
Calculated	Calculated			
Scores	Scores	Mean Difference	Std. Error	Sig.
CT	SS	-.252	.034	.000
	RI	-.456	.035	.000
SS	RI	-.204	.026	.000

I rejected the null hypothesis at the .05 level. I found statistically significant differences between the means of all three total calculated scores (Table 9). Specifically, the CT mean score (2.676) was significantly lower than all of the other scores. While higher than the CT mean score, the SS mean score (2.929) was lower than the RI score (3.132).

Factor Analysis

Ordinarily, a factor analysis of items in an instrument is completed before scale reliability analyses are done. However, in this study, because I was using Vinson's previously developed instrument, and that instrument was based on a theory of five orientations, and each orientation had a related scale of six items, I did the scale analysis first on those intended scales. Leaving my intentions aside as to what the scales should theoretically measure, a principle component factor analysis with both oblimin and varimax rotation was conducted to determine which items held together in a predictable manner. After determining that enough items existed across the five scales to validly conduct a factor analysis, multiple versions of factor analysis were conducted with items

forced into five factors and then into three factors in an effort to find a factor solution that showed individual items loading strongly and clearly onto one factor and factors showing theoretically appropriate loadings of items.

Analysis of output generated through the principle component factor analysis with varimax rotation proved unfruitful with multiple items loading weakly across multiple components with little connection to current theory about teaching orientations in history and social studies. However, an examination of the pattern matrix generated from a principle component factor analysis with direct oblimin rotation where items were forced into three factors showed the greatest promise. After removing four items of the 30-item inventory, an examination of the pattern matrix generated from the factor analysis with direct oblimin rotation showed that each item was loading strongly onto only one component and the three components had clear theoretical implications (Table 10). For this three-factor solution, The Kaiser-Meyer-Olkin Measure of Sampling Adequacy was .802 and Bartlett's Test of Sphericity was significant at the .000 level. Each of the three components displayed Eigenvalues over 2. Together, the three components explained close to 41% of the total variance in items. This three-factor solution was chosen to analyze and describe.

Table 10: *Three-Factor Forced Principle Component Analysis with Direct Oblimin Rotation*

Component	Item	Load Value
Component One	07. Students should methodically analyze social problems	.752
	08. Students should view American institutions critically	.633
	10. Students should identify and work through problems with processes outlined by social scientists	.577
	13. Students should defend methods of social science investigation	.468
	14. Students should demonstrate understanding of how to succeed personally & academically	.508
	15. Students should methodically inquire into problems relevant to experiences	.637
	17. Teachers should discuss power inequality	.440
	20. Teachers should teach methods of inquiry as a means to personal fulfillment	.533
	21. Teachers should teach research skills of professional social scientists	.648
	22. Teachers should teach students to practice skills of democratic activism	.609
25. Teachers should use as many instructional strategies as possible to maximize the number of students who are successful	.413	

Component	12. Students should possess positive self esteem	-.544
Two	24. Teachers should utilize methods that increase students' self esteem	-.597
	27. Content should emphasize minority groups	-.574
	29. Content should be selected based upon maximizing individual students success	-.695
	31. Content should be based upon student-identified problems	-.786
	33. Content should arise from investigating problems derived from students needs and interests	-.763
	34. Content should focus on past & contemporary examples of racism, sexism, elitism	-.687
	35. Content should student-selected	-.790
Component	08. Students should view American institutions critically	-.350
Three	16. Students should believe that democracy is best form of government	.524
	19. Teachers should lecture	.614
	26. Teachers should assess with objective multiple choice questions	.689
	28. Content should be standardized	.451
	36. Content should come from textbooks	.641

The first component generated by the factor analysis contained items that were intended originally to measure the preferred purpose and pedagogy of RI-oriented teachers. The second component contained items that described who should make content choices in the social studies. The final component contained items intended to measure the CT orientation. The nature of the collection of items under each component led to a few conclusions about what each component is theoretically measuring. First, if one were to assume that there are five teaching orientations in HSS (as I posit at the outset of this study), and that the items that were intended to measure the five orientations were effective, those items would load together, predictably and strongly, onto the appropriate component of the five. This was not the case. Therefore, the five orientations thesis is not the best way to understand Massachusetts HSS teachers.

Second, if one were to assume as Barr, Barth and Shermis do, that there are in fact three traditions in HSS, the items intended to measure each of those three traditions would load together, predictably and strongly onto the same component. Once again, this was not the case. In fact, items that were intended to measure the pedagogical and philosophical priorities of Barr, Barth and Shermis' second orientation, the SS orientation, loaded strongly and unpredictably into the first component along with over ten RI items. Therefore, in light of the factor analysis, the traditional methods used previously to measure the three orientations are not the most effective way to understand the HSS teaching sample in Massachusetts, owing to the fact that the SS orientation items are not differentiable from RI orientation items in the factor analysis. In the next section, I suggest a theory as to why these items did not load into their own components.

The items that appear together in the first component and the second component do, however, have an underlying theoretical link. It appears the differentiating themes among the components are what one believes about why and how HSS is taught on the one hand and what is taught on the other. The first component, which is a scale that measures “why and how” one teaches HSS, loaded items that were only meant to appeal to RI orientation teachers. It may well be that someone whose calculated orientation is CT would have a low scale score with these items – in other words, they may disagree or strongly disagree that these items represent how and why HSS teachers teach their classes. Conversely, it may be that someone whose calculated orientation is RI, RIPE or RISE would have a high scale score with these items. In the second component where “what” one teaches in HSS is addressed, items that were meant to get at the RIPE orientation loaded strongly.

Again, it may well be that someone with a calculated CT orientation would rate the notion of student-chosen content for the purpose of “increasing students’ self esteem” unfavorably while someone with an RI, RIPE or RISE calculated orientation would rate those items favorably. The final third component had six items loading strongly on it – five of the six items were items intended to measure the CT orientation. The sixth item on this component was an opposite reaction to a RISE item claiming that a purpose of HSS education should be to help students “view American institutions critically.” One could theorize that individuals whose calculated orientation is CT would score highly on this scale while one whose calculated or selected orientation is RI would score lower on this

scale. The fact that the last item loads negatively is in keeping with the CT orientation because, hypothetically, these teachers do not view American institutions critically.

One-way ANOVA tests were conducted to determine whether there were statistically significant differences in scores for teachers from different calculated orientations on the first and second component. The expectation was that someone from CT orientation would score significantly lower on both scales while someone from the SS orientation would score significantly higher than the CT orientation and significantly lower than the RI orientation. Individuals from the RI orientation were expected to have the highest mean scores for the three groups on both of these scales.

$$H_0 : \mu_{CT} = \mu_{SS} = \mu_{RI}$$

$$H_A : \mu_{CT} < \mu_{SS} \text{ OR}$$

$$H_A : \mu_{CT} < \mu_{RI} \text{ OR}$$

$$H_A : \mu_{SS} < \mu_{RI}$$

Table 11: *One-way ANOVA Test of Differences in Scores on Component One Based on Teachers' Calculated Orientations*

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1.435	2	.718	6.416	.002
Within Groups	24.607	220	.112		
Total	26.042	222			

Table 12: *One-way ANOVA Test of Differences in Scores on Component Two Based on Teachers' Calculated Orientations*

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	16.440	2	8.220	42.027	.000
Within Groups	43.411	222	.196		
Total	59.862	224			

Both one-way ANOVA tests showed statistically significant differences at the .05 level between teachers' scores based on their different calculated orientations (Table 11 and 12). Post-hoc analyses using the Scheffe test were conducted to determine where the statistical significance(s) in mean scores appeared. The Scheffe test for scores on component one showed that the mean score for reflective inquiry teachers ($M = 3.42$, $SD = .344$) was significantly higher than cultural transmission teachers ($M = 3.02$, $SD = .186$). Social science teachers ($M = 3.33$, $SD = .286$) did not differ significantly from either RI or CT teachers. The Scheffe test for scores on component two indicated that the mean score for reflective inquiry teachers ($M = 3.02$, $SD = .431$) was significantly higher than both social science teachers ($M = 2.44$, $SD = .493$) and cultural transmission teachers ($M = 1.99$, $SD = .529$). Social science teachers scored significantly higher than cultural transmission teachers on this component. In other words, with only one exception, the One-way ANOVA tests showed statistically significant differences in scores on component one and two in precisely the directions predicted.

Once the principle component factor analysis was taken into consideration, scale reliability analyses were run again based on the items loading on each of the three factors. Two of the three factors showed a strong Cronbach's Alpha. Specifically, scaling the items from the first component yielded a Cronbach's $\alpha = .814$. Scaling the items from the second component yielded a Cronbach's $\alpha = .863$. Unfortunately, even though the factor analysis showed a third component which loaded all CT items (thus, making me hopeful that I could perfect a CT scale for this study) the scale reliability analysis of these CT items in the third component yielded an unsatisfactory Cronbach's $\alpha = .575$. The results from the Principle Component Factor Analysis therefore point to two reliable scales: 1) Items used together to measure teachers' pedagogical and philosophical approaches to HSS and 2) Items used together to measure teachers' beliefs about the role of students in making content decisions. These two scales do not measure directly the three orientations; however participants from different orientations score significantly differently from one another in predictable directions. This will also be discussed in the next section.

Comparing the performance of this instrument against a similar instrument used by Vinson for his 1997 dissertation and a subsequent publication (Vinson, 1996, 1998) shows that this instrument appears to perform better, the scales and subscales perform more predictably and reliably. Vinson's results from the original administration of a similar instrument contained a few flaws that were quickly remedied. For example, rather than conducting scale reliability analyses on the scales and subscales and deriving Cronbach's α , Vinson apparently conducted correlation analyses to determine if there

were statistically significant correlations between items intended to measure the same thing. Vinson's results showed that some of the items intended to measure the same construct were not significantly correlated. Some items that were significantly correlated were actually negatively correlated. An examination of Vinson's survey showed that these items had several conflicting theoretical messages. Namely, items intended to measure the RI tradition which, at their core should focus on the students making choices about content and student-focused pedagogical approaches, actually had stems referring to what "teachers should" do. Still, other items were double, even triple barreled or had awkward phrasing. Vinson acknowledged some of these issues in the body of his concluding chapters. For this study, I attempted to amend these oversights and I corrected awkward statements. The fact that the modified version of the Vinson instrument that I administered in early Spring 2009 had scales and subscales that were at least reliably measuring the same thing appears to be an improvement over the Vinson model.

If this survey were to be re-administered, researchers should continue to try to perfect the items that, in theory measure these three different orientations. In addition, future researchers should try to replicate the finding (generated by the factor analysis in this study) that two additional scales can serve to differentiate HSS teachers. Specifically, they should test the reliability of a scale of items that measures the purpose/pedagogy preferences of HSS teachers, and a second scale that measures the preferred content of HSS teachers (see Table 13).

Table 13: *Items Included in Two New Scales to Assess Curricular Orientation*

New Scale	Items included in new scale	Cronbach's
		α
Preferred	07. Students should methodically analyze social problems	.814
Purpose and	08. Students should view American institutions critically	
Pedagogy	10. Students should identify and work through problems	
Scale	with processes outlined by social scientists	
	13. Students should defend methods of social science	
	investigation	
	14. Students should demonstrate understanding of how to	
	succeed personally & academically	
	15. Students should methodically inquire into problems	
	relevant to experiences	
	17. Teachers should discuss power inequality	
	20. Teachers should teach methods of inquiry as a means to	
	personal fulfillment	
	21. Teachers should teach research skills of professional	
	social scientists	
	22. Teachers should teach students to practice skills of	
	democratic activism	
	25. Teachers should use as many instructional strategies as	
	possible to maximize the number of students who are	
	successful	

Preferred	12. Students should possess positive self esteem	.863
Content Scale	24. Teachers should utilize methods that increase students' self esteem	
	27. Content should emphasize minority groups	
	29. Content should be selected based upon maximizing individual students success	
	31. Content should be based upon student-identified problems	
	33. Content should arise from investigating problems derived from students needs and interests	
	34. Content should focus on past & contemporary examples of racism, sexism, elitism	

Participants' calculated orientations can be used in conjunction with their scores from scales presented in Table 13 to improve items intended (in theory) to measure the CT orientation. Future researchers should expect to see on these first two scales that CT oriented teachers score low on the scale and teachers working from an RI orientation score high on the scale. Cut points along this field to delineate one orientation from the next should be set in future studies. Generally, the results of the factor analysis could be used to target and highlight items that create the greatest level of differentiation in responses. This could reduce the number of items needed to survey orientations by eliminating items that do not create strong differentiation among the sample group or show up as low loading items in future factor analyses.

Measuring and Understanding Teachers' HSS Orientations

By way of a final analysis regarding this instrument's ability to measure and distinguish between orientations of HSS teachers, it is clear that there are some significant limitations to studies that try to understand the HSS teaching population in terms of five orientations framework, three orientations framework or even two orientations. Rather than trying to understand individual members of the HSS teaching population as belonging "in" one orientation or another, it appears that it is more appropriate to understand teachers' individual orientations as points in a quadrant on a field rather than as contained within inflexible and discrete categories. By visualizing the orientations as quadrants on a field, much like a tennis court, it follows that any two individuals who are categorized within one quadrant are not necessarily qualitatively identical. As explained in detail later in this section, teachers within the same quadrant can differ quite a lot from one another. Using the tennis court metaphor, several teachers may be in the same quadrant but still quite far apart from one another. This interpretation allows for significant differences between teachers who, on the surface, appear to be the same in light of the fact that they are within the same orientation. Orientations that have been theorized in previous studies – and the often contradictory blends between and among those (supposedly) discrete orientations - make greater sense when the orientations are overlaid onto a field map rather than treated as detached and isolated categories (see Figure 2).

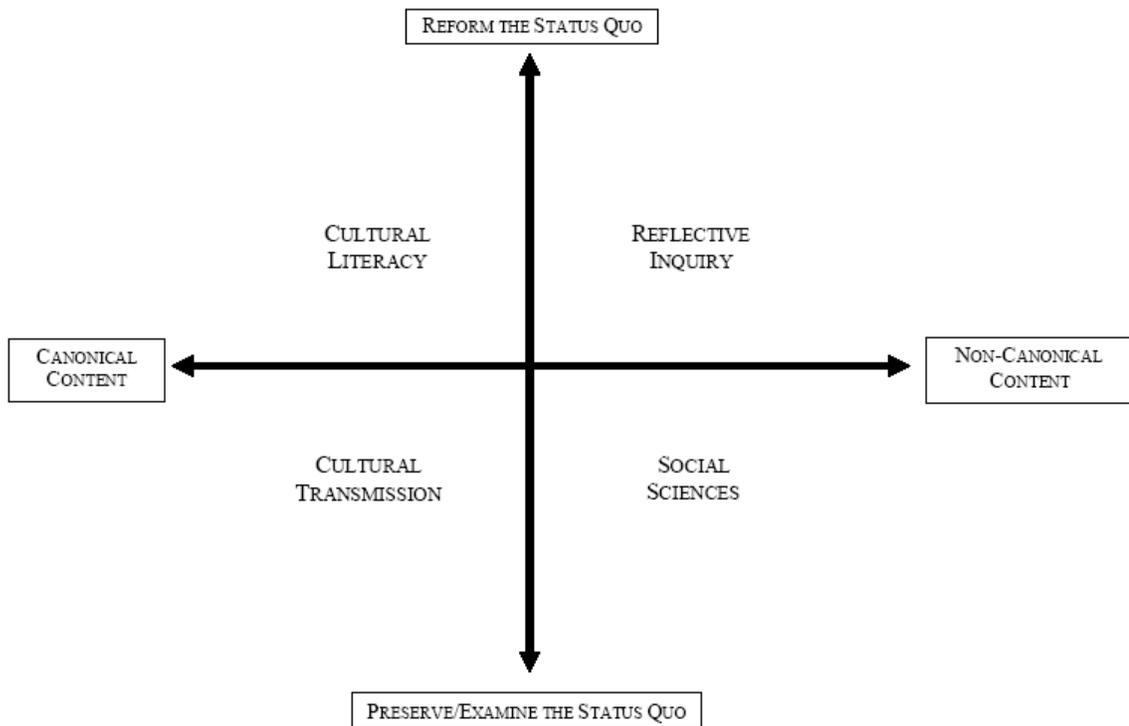
In this field map, orientations are measured along two continua. The y-axis represents a continuum of the messages that teachers endorse about either reforming or

conserving society. Items that loaded together in the first component generated by the principle component factor analysis (Table 13) can be used in future studies to measure where along this continuum individual teachers fall. The x-axis represents a continuum of content choices that teachers within the orientation make spanning from the more conservative (content choices based on expert and/or authoritative input) to the more liberal (content choices based on student's interests or non-canonical history such as social problems based curriculum). Items that loaded together on the second component generated by the factor analysis (Table 13) can be used to measure where along the continuum individual teachers fall. I connect this field map of orientations to the findings from this survey in the next section.

The CT and SS orientations are below the horizontal axis owing to their more conservative messages about society and the status quo – that is to say that teachers within these orientations do not seek to reform the status quo. Meanwhile, the RI orientation and what I call the “Cultural Literacy” quadrant are above the horizontal axis owing to their more liberal messages about reforming society. The Cultural Literacy quadrant has not, heretofore, been acknowledged or discussed in the literature on teaching orientations. However, using these concepts to frame the axes (non/canonical content choices on the x-axis, and preserving or reforming the status quo on the y-axis) leaves a vacuum in the top left quadrant. I borrowed Hirsch's (1988) term Cultural Literacy to label and describe this quadrant due to the fact that teachers who work within this top left quadrant support the use of canonical content to reform society. This quadrant appears merely as a theoretical addition to the conversation on HSS teaching

orientations and is discussed further below. Teachers' orientations therefore can be understood as predominately belonging in one of the four orientations, or quadrants. However, within that quadrant, teachers present much more nuanced images of that orientation.

Figure 2: *Field Map of Orientations*



For example, the social science orientation has been theorized as a culturally conservative orientation in that teachers who subscribe to this orientation tend to use social science inquiry methods to examine the world as it is. For this reason, the social science orientation is found in the bottom right hand corner of the field map where liberal content

choices intersect with more conservative or moderate messages about the status quo. In the sense that the content explored in the SS teacher's classroom is non-canonical (i.e. social problems rather than traditional content), some SS teachers may be more closely aligned with the culturally liberal reflective inquirers rather than the culturally conservative cultural transmissionists. They would place closer to the top of the SS quadrant nearer to the RI quadrant. Conversely, some SS teachers might align more with cultural transmissionists in that neither CT nor SS oriented teachers are interested in encouraging their students to reform society and both look to authorities such as social science experts to generate content ideas. These teachers, rather than being at the center of the SS quadrant are more likely to be found closer to the left edge of the quadrant nearer to the CT quadrant.

For another example, some teachers who score highest in the reflective inquiry orientation may be interested in having students explore social problems and generate their own solutions to those problems. This reflective inquiry approach toward content may be overshadowed by a more moderate approach to social reform; the teacher may be interested in individual students' personal development rather than educating future agents of social change. While the content choices of these teachers may be inquiry oriented in some respects, these teachers' aims are far more moderate than the aim of large scale social reform, which those further inside the quadrant would advocate. Rather than being in the center or top right of the teachers within the reflective inquiry quadrant, these teachers are more likely to be found toward the bottom edge of the quadrant and closer to the more conservatively oriented teachers working from a CT or SS orientation.

Understanding HSS orientations as points within a field also highlights a group that has received little if any attention from researchers who theorize orientations to HSS teaching. Only one person in this study showed a tie in their highest calculated scale score between CT and RISE orientations. According education researchers Barr, Barth and Shermis (1977; Barr, Barth, & Shermis, 1978), this teacher is theoretically confused – one cannot subscribe on the one hand to a liberal purpose and pedagogy but make culturally conservative content choices. However, a few prominent education policy researchers (Hirsch, 1996; Ravitch, 2000; Thernstrom & Thernstrom, 2000) claim to promote a liberal social reform agenda (i.e. closing the racial achievement gap and creating greater equity in schooling), but simultaneously they endorse a very conservative approach to content choices. These educators are likely to be found in the top left quadrant of the field image (Figure 2). Understanding the orientations of HSS teachers as points in a field allows for a more holistic interpretation of history and social studies teachers where overlaps between orientations are both expected and theoretically sound.

Additionally, understanding HSS teaching orientations in this manner accounts for the variance observed across empirical studies that try to assess concentrations of orientations during any given time and in any given place. For example, when looking at a national sample, one might expect to find a relatively normal distribution of HSS teachers among each of the four quadrants or within each orientation in the field. However, any given sample of the teaching population is likely to be skewed and show peaks and valleys in different quadrants within this field. For example, Anderson, et al (1997) linked particular teaching orientations to demographic characteristics including

political ideology and geographic location. Teachers living in the southern United States, and those from small cities and towns, they argued, tended to identify with more conservative teaching approaches and self-identified as Republican. Meanwhile, those living in the Pacific states or the northeastern states tended to align with more liberal teaching orientations and identified themselves significantly more often as Democrats. By viewing orientations as peaks in frequencies within this field, it is possible that three orientations are in ascendency – or the most subscribed to orientations by HSS teachers – in one place while one of those orientations is eclipsed entirely in another geographic location by another curricular orientation. Such was the case in this study of Massachusetts teachers where a very small minority of teachers was found to be cultural transmissionists. This will be discussed more in the next section. By viewing HSS teaching orientations in this manner, it is clear that the variance in observations of teachers’ HSS orientations between different empirical studies is likely an effect of sampling rather than indicating large pendulum shifts in, or the changeability of, the HSS teaching force.

Using the field map of orientations can be useful in the interpretation of survey data – it can make greater sense of the original three orientations by showing how teachers can have leanings toward more than one orientation simultaneously. A field map interpretation of HSS orientations highlights the complexity with which teachers understand and approach their work. However, the categories “cultural transmissionist,” “social scientist” and “reflective inquirer” and the items used to measure the prevalence of these orientations is still useful for researchers. Continuing to use these categories to

understand concentrations of orientations in the HSS teaching population will suit many purposes. For example, if one needed to make policy decisions based on what HSS teachers of any given geographic area saw as central to their work, the survey instrument and items would prove germane to that task.

Based on the factor analysis, t-test, and scale reliability analyses, the sample of Massachusetts HSS teachers studied for this dissertation is best represented by the three orientations thesis rather than the proposed five orientations alternative. As noted, both the t-test and the scale reliability analysis showed that the two orientations in the “RI family” – the RI and RIPE orientations – were not differentiated from one another. This suggests that distinguishing between the RI, RISE and RIPE orientations is not useful. Rather, one can understand the different leanings within the RI category (i.e. the “personal ends” or “social ends” leaning) as an indication of that teachers’ placement within the RI quadrant. Again, the RIPE category, though in the RI quadrant, would be toward the lower border of the quadrant. The RISE, while in the RI quadrant, would be toward the upper border of the quadrant. Regardless, the same quadrant is used for the two different types of teachers. For that reason, the three orientations are used for the remainder of this chapter as a way to summarize findings about Massachusetts teachers. The three orientations are also used as a frame for the content analysis of the Massachusetts Comprehensive Assessment System (MCAS) discussed in the next chapter. Using the three orientations framework in both data analyses created a basis of comparison between the two different sets of data collected. This common basis for

analysis is crucial to the second part of chapter five in which I explore the relationship between HSS teachers in Massachusetts and the MCAS test in United States history.

The Prevalence of the Revolutionary Spirited RI Orientation in Massachusetts

The findings of this study show that orientations of Massachusetts HSS teachers are unlike those of teachers studied in earlier research. In the following section, I make the case that Massachusetts teachers appear to be more critical in their teaching approach and more liberal in their political ideology than HSS teachers nationwide. Apparently, even transmission oriented teachers in Massachusetts are more centrist leaning than their fellow CT oriented teachers across the country. I first discuss the proportions of teachers who align with each of the three orientations and how, specifically, the Massachusetts HSS teachers surveyed differ in significant ways from the national HSS teaching population. Next, I account for some of the differences between the Massachusetts HSS teachers and the national population by highlighting some key demographic differences between the two groups. Finally, I superimpose the findings about Massachusetts HSS teachers onto the “Field Map of Orientations” and consider the findings in light of the theoretical frame.

Reformers and Reflective Inquirers: Massachusetts HSS Teachers

Massachusetts residents have long battled injustice on many fronts. From its revolutionary inception as the “Cradle of Liberty,” where colonists seasoned the waters of Boston Harbor with hundreds of chests of unreasonably taxed tea, to the battlefield at Lexington and Concord where Minutemen fired the “shot heard round the world,” Massachusetts residents have long fought the good fight for equality and justice.

Massachusetts was the first state to offer free public education to its citizens, and lead by its native son, William Lloyd Garrison, it was the first state to abolish slavery in 1783 (though Pennsylvanians contest that claim). During the Civil War, Massachusetts was one of the first states to establish a free, all-Black regiment. At about the same time, Massachusetts women were playing a key role in the movement for women's suffrage. In the 1980s, Massachusetts was among the first to send an openly gay Congressman, Barney Frank, to the House of Representatives. And, in 2004, Massachusetts became the first state to legally recognize same sex marriage. Massachusetts' long reformist roots are alive and well within its residents, even after George H.W. Bush's 1988 general election campaign made repeated lowbrow attempts to isolate Massachusetts citizens from the rest of the nation by referring to them derogatorily as "those Massachusetts liberals." Indeed, those reformist leanings seem to be present in the orientations of Massachusetts history and social studies teachers.

Overwhelmingly, the survey responses of Massachusetts HSS teachers indicated social reformist orientations. These teachers responded strongly and favorably to the notion that oppressed and/or minority groups in American (i.e. women, people of color, the laboring class) should be highlighted in history and social studies curriculum. They claimed to lead discussions with their students about the roots of inequality and its current manifestations in the United States and abroad. They not only discussed inequality, but also asked that their students *combat* inequality by participating on the democratic stage and actively working for social justice. According to the survey, for the majority of Massachusetts HSS teachers, the history and social studies classroom is not

merely a place to transmit historical wisdom to students. Nor is it a place where adolescents, like budding scientists, should examine society as it is to make empirical judgments about how it came to be that way. Rather, the HSS classroom in Massachusetts is a place where authentic and troubling social problems are discussed, the roots of the issues are ferreted out, and students as young democratic participants are asked to reason through the issues, and to fight for increased social justice by targeting the cause of the problem. While this may not be the scene in *every* HSS classroom in Massachusetts, teachers in this study overwhelmingly demonstrated through their responses to survey questions that this is the idealized version of their classroom.

Massachusetts teachers' beliefs about the purpose, method and content of history and social studies (or teachers' "orientations") were determined several different ways. First, teachers' total scores on each of the scales, representing each of the theorized orientations, were calculated. Teachers' calculated mean scores for each orientation were computed by dividing the total scale score, or raw score by six. Next, teachers' *Calculated Orientations* were assigned based on their highest total raw score across the orientations total raw scores. Teachers who skipped one item within any of the scales meant to measure teaching orientations were treated by averaging their responses to the five remaining items in the scale and then using that as their scale scores. Scale scores for these individuals were examined and their calculated orientations were assigned as their highest scale score between the five scales. Others who had more than one item missing from any one of the five scales meant to measure calculated orientation were designated as "missing data" for the calculated orientation. This reduced the number of teachers with

calculated orientations from the full number of participants (n=272) to (n=257). Chi-Squared tests of association were conducted to determine whether there were any differences between the demographic characteristics of teachers who had a calculated teaching orientation and those who were designated as missing data owing to skipping an item with the scales. The output generated from these tests showed no significant differences between those with calculated orientations and those with missing data in terms of their gender, selected teaching orientation, political identification or number of years teaching. That is to say, the group of teachers who had missing data for their calculated orientation did not differ significantly from the group who had calculated orientations in terms of those demographic characteristics. There is, therefore, reason to conclude that the results of analyses that were dependent upon a teachers calculated orientation were not skewed by omitting those with missing data.

Assuming five orientations as I theorized at the outset of this study, the calculated orientations are reported on Table 14. Although I began the analysis with five orientations, the RIPE and RISE category were eventually subsumed under one RI category. I report the five categories here to show that while the three orientations subsumed under the RI category were not statistically different, teachers showed the most favorability toward RISE items. As I argue throughout the remainder of this chapter, Massachusetts HSS teachers' predilection toward the RISE items indicates their strong support of using inquiry methods to reform society – a strong indicator that Massachusetts teachers are unlike teachers nationwide in their reformist tendencies.

Table 14: *Calculated Orientation Frequencies Based on Five Orientations Framework*

Curricular Orientation	Frequency	Percent
Cultural Transmission (CT)	11	4.3
Social Sciences (SS)	24	9.3
Reflective Inquiry (RI)	25	9.7
RI for Social Ends (RISE)	125	48.6
RI for Personal Ends (RIPE)	24	9.3
Tie between CT & RISE	1	.4
Tie between CT & RIPE	1	.4
Tie between SS & RISE	5	1.9
Tie between SS & RIPE	1	.4
Tie between RI & RISE	11	4.3
Tie between RI & RIPE	5	1.9
Tie between RISE & RIPE	10	3.9
Tie among three or more	14	5.4
Total	257	100

In this sample, 81.3% of teachers had one score that was clearly the highest of the orientations (n = 209). However, 13.3% of teachers in this sample had ties for their highest total scores between two orientations (n = 34) and 5.5% of teachers in this sample had ties for their highest total scores among three or more orientations (n=14). In other

words, most Massachusetts HSS teachers in this sample had very clear inclinations toward a singular orientation.

Using the Five Traditions as a framework, it is clear that the reflective inquiry for social ends orientation dominates among Massachusetts HSS teachers in this sample with 48.6% (n=125) having RISE as their highest scale score. What this means is that a significant percentage of Massachusetts HSS teachers explicitly condoned using their history and social studies classroom as a site for examining critical social problems and investigating, or even instigating, reform efforts. For example, many positively responded to items that claimed that as a result of history and social studies, students should be able to “view American institutions and traditions with a critical eye” (Item 8) and “work toward social justice and equality” (Item 11).

The next highest percentage of teachers were reflective inquirers with 9.7% (n = 25) having RI as their calculated orientation, and reflective inquiry for personal ends claimed the next highest percentage with 9.3% of teachers (n = 24) having RIPE as their calculated orientation. These teachers, while a little less likely to encourage their students to agitate for social reform, nonetheless believed that their students should be qualifiedly affected or altered by the history and social studies classroom experience – in other words, one of their primary goals was reform of the individual. For example, these teachers responded positively to an item asserting that as a result of history and social studies, students should “demonstrate increased understanding of how they can succeed personally and academically” (Item 14) and another item that claimed that history and

social studies content should “be selected based upon maximizing student success and growth” (Item 29).

In light of the fact that previous statistical analyses demonstrated that the three orientations found in the RI family work better together than alone, this table of calculated orientations was collapsed into three orientations. The three lines in 14 that show ties between any one of the three orientations in the RI family, along with the three lines showing RI, RISE and RIPE individually were subsumed under one category called RI orientation. In that case, 77.7% of teachers (n = 200), had their highest score in the three orientations associated with the reflective inquiry tradition. Meanwhile, only 4.3% of this sample have their highest score in the CT orientation (n = 11). This orientation to HSS teaching reflects much more traditional notions of what should be taught and how it should be taught. The few teachers in this sample characterized as teachers working from a CT orientation responded strongly and favorably to items asserting that, as a result of history and social studies, students should “believe that democracy is the best form of government” (Item 16) and that lecturing is a preferred pedagogical approach (Item 19). They also agree that content should be standardized across all districts (Item 28) and that content should come mostly from textbooks (Item 36).

A larger number of teachers (n = 24, 9.3%) had their highest score in the SS orientation. While not as traditional as the CT oriented teachers are, SS oriented teachers endorse the authoritative role of social scientists to define the content and curriculum (Item 30). They also agreed with items suggesting that rather than focusing class activities on discussion and inquiry, as RI oriented teachers do, HSS teachers should

focus specifically on teaching the research skills that professional social scientists use (Item 21) and investigating the problems that professional social scientists have identified (Item 23). The purpose of these activities is not to solve social problems but rather simply to identify and work through problems using the processes that social scientists use (Item 10). The fact that the vast majority of teachers in this sample align with the RI teaching orientation and only a small minority align with the CT or SS orientations demonstrates, in sharp relief, that the majority of Massachusetts HSS teachers take as one of their primary mandates either the reform of individual students and/or the reform of American society and institutions.

There was also a clear link between teachers' calculated orientations and their political identities. Recall that White (1982) made the contention that there appeared to be two orientations: a conservative CT-type orientation, and a liberal RI-type orientation. Meanwhile Hirsch (1996) contended that a teacher's approach to teaching, whether transmission oriented or critically oriented, had little to do with a teacher's political identification. This sample of Massachusetts HSS teachers were asked, "Which of the following best describes you?" with the choices listed as "very liberal," "liberal," "moderate," "conservative," and "very conservative." I used a Chi-Squared test of association to determine if their answers to this question were related to their calculated teaching orientation (Table 15).

Table 15: *Chi-Squared Test of Association between Calculated Orientation and Political Identification*

Calculated orientation	Political identification			X ²
	Liberal	Moderate	Conservative	
RI	112 (.8)	76 (-.4)	10 (-1.2)	14.82*
SS/CT	10 (-1.9)	18 (1.0)	7 (2.8)	

Note. * = $p < .05$. Adjusted standard residuals appear in the parentheses below group frequencies.

The findings shed doubt on Hirsch’s contention of no relationship between teaching orientation and political identification. Teachers in this sample who had calculated cultural transmission or social science orientations tended to identify themselves as more politically conservative at statistically significantly higher frequencies than expected ($p < .05$). Contrary to Hirsch’s contention, there was an association between teaching orientation and political ideology. This link has particular bearing in the analysis of the demographic characteristics of this sample versus the nationwide HSS teaching population.

Massachusetts teachers in this sample had a relatively uniform idea about what HSS curriculum and instruction should look like as indicated by their reflective inquiry approach to history and social studies education. In theory, this orientation is strikingly similar to those who work from a democratic pragmatist lens and who advocate for

particular strategies to prepare youth for active citizenship in a democracy. These implications will be dealt with in greater depth throughout the last three sections.

True Blue Teachers in a Blue State: Demographically Distinct from US teachers

In the book, *The Bluest State*, author John Keller (2007) refers to Massachusetts as “a Democrat’s Burger King: They always have it their way” (Jacoby, 2009). It comes as no surprise then that Massachusetts HSS teachers reflect the strong slant that citizens of this small, but powerful, state embody writ large. In this section, I discuss the demographic characteristics of HSS teachers who work in “The Bluest State” and compare these teachers with the characteristics of secondary history and social studies teachers nationwide.

Leming’s (1991) comprehensive review of HSS teachers’ characteristics was used as a baseline of comparison for my sample of social studies teachers. Leming’s study of teachers’ characteristics was a quantitative review of several empirical studies including surveys conducted by the National Council for the Social Studies (NCSS), the National Center for Education Information [NCEI (SS)] and High School and Beyond (HSB). As stated previously, although Leming’s review was based on surveys from as far back as 20 years ago, it remains the most current and comprehensive review to date of the demographic characteristics of HSS teachers. The demographic findings of these studies were compared to the findings of the current study using Chi-Square goodness of fit statistical tests. Demographically speaking, the respondents were similar to the history and social studies teaching population measured in previous studies with some notable differences (Table 16).

Table 16: *Racial Demographic Comparisons between Massachusetts and Nationwide HSS Teachers*

Racial Identification	Massachusetts	Nationwide
White	83.4%	92%
African American	1.9%	1.2%
Asian American	1.9%	-- ^a
Hispanic American	0.4%	--

Note. Nationwide figures were obtained from Leming, J. S. (1991). Teacher characteristics and social studies education. In J. P. Shaver (Ed.), *Handbook of research on social studies teaching and learning* (pp. 661). New York: McMillan Publishing Company.

^a. Dashes indicate that data were not reported for the cell.

Among the Massachusetts HSS teachers surveyed for this study, an overwhelming majority of respondents identified themselves as White (93.4%, n=239). Less than two percent identified themselves as African American (1.9%) or Asian American (1.9%). Meanwhile, Leming reported that data gathered via surveys conducted by NCSS, NCEI (SS) and (HSB) showed that 92% of social studies teachers were White. The NCSS survey found that 1.2 % of teachers identified themselves as African American. Results of the Chi-Squared goodness of fit test (Table 17) yielded a Chi-Square value of 5.975. Meanwhile, the critical Chi-Square value at the .05 significance level is 5.99.

Table 17: *Chi-Squared Goodness of Fit Test between Massachusetts and Nationwide HSS Teachers in Terms of Race*

	Observed		Expected %		$(P_j - 100\pi_j)^2$	$\frac{(P_j - 100\pi_j)^2}{100\pi_j}$
	f_j	P_j	$100\pi_j$	$P_j - 100\pi_j$		
White	239	93.4	92	1.4	1.96	0.02
Non-White	9	3.6	8	-4.4	19.1	2.4
					Total = 2.4	
					$f_j / 100 = 2.48$	
					$\chi^2 = 5.98$	
					Crit χ^2 (p=.05) = 5.99	

Because the observed chi-square value was less than the critical value at the .05 level of significance, there is evidence of goodness of fit between the proportions of races observed in the current sample of Massachusetts teachers and the expected proportions based on Leming's study. In other words, in this sample, Massachusetts HSS teachers' racial characteristics are statistically similar to those proportions measured in the nationwide social studies teaching force previously reported.

Second, the majority of survey respondents identified themselves as male (63.1%, n=152) while a sizable minority identified themselves as female (36.1%, n=87). Again, this is similar to trends found in Leming’s 1991 review.

Table 18: *Gender Demographic Comparisons between Massachusetts and Nationwide HSS Teachers*

Gender Identification	Massachusetts	Nationwide
Female	36.1%	25%
Male	63.1%	75%

Note. Nationwide figures were obtained from Leming, J. S. (1991). Teacher characteristics and social studies education. In J. P. Shaver (Ed.), *Handbook of research on social studies teaching and learning* (pp. 661). New York: McMillan Publishing Company.

He wrote (1991) that while the vast majority of *teachers* are female, the ratio is not as skewed toward female teachers at the secondary level. In fact, the ratio is reversed for *social studies teachers* at the secondary level where male teachers outnumber female teachers by a ratio of 3:1 (226).

Table 19: *Chi-Squared Goodness of Fit Test between Massachusetts and Nationwide HSS Teachers in Terms of Gender*

	Observed		Expected %			
	f_j	P_j	$100\pi_j$	$P_j - 100\pi_j$	$(P_j - 100\pi_j)^2$	$\frac{(P_j - 100\pi_j)^2}{100\pi_j}$
Male	152	63.6	75	-11.4	130	1.73
Female	87	36.4	25	11.4	130	5.2
Total = 6.93						
$f_j / 100 = 2.39$						
$\chi^2 = 16.57$						
Crit χ^2 (p=.05) = 3.84						

However, results of the Chi-Squared goodness of fit test (Table 19) yielded a Chi-Square value of 16.57. Meanwhile, the critical Chi-Square value at the .05 significance level is 3.84. Because the observed chi-square value exceeded the critical value at the .05 level of significance, there was evidence of lack of fit between the proportions of gender observed in the current sample of Massachusetts teachers and the expected proportions based on Leming’s study due to the greater presence of females in this sample of Massachusetts HSS teachers than expected. Regardless of the evidence of lack of fit between the two samples’ gender ratios, Leming’s (1991) contention that, “at the secondary level, the social studies profession has a substantially higher percentage of male teachers than the profession as a whole” (226) continues to ring true for this study. Male teachers did

outnumber female teachers in this sample of Massachusetts secondary level HSS classrooms by a ratio of 1.75 to 1.

Additional notable differences were found between the demographic characteristics of teachers in the current study and those found by Leming. Specifically, teachers in this study were much younger, less experienced and far more politically liberal than social studies teachers found in Leming's study.

Table 20: *Age Comparisons between Massachusetts and Nationwide HSS Teachers*

Age	Massachusetts	Nationwide
30 years or younger	27.3%	10%
31 – 40 years	33.5%	32%
40 +	39.2%	58%

Note. Nationwide figures were obtained from Leming, J. S. (1991). Teacher characteristics and social studies education. In J. P. Shaver (Ed.), *Handbook of research on social studies teaching and learning* (pp. 661). New York: McMillan Publishing Company.

The majority of respondents in the current study (65.5%) responded they had taught less than 12 years. Massachusetts HSS teachers are correspondingly young with 27.3% (n=71) between the ages of 20 – 30 years, and 33.5% (n=87) between the ages of 31 – 40. The data suggest that the Massachusetts HSS teaching force is much younger than the HSS teaching force that Leming described in 1991.

Table 21: *Chi-Squared Goodness of Fit Test between Massachusetts and Nationwide HSS Teachers in Terms of Age*

	Observed		Expected %			
	f_j	P_j	$100\pi_j$	$P_j - 100\pi_j$	$(P_j - 100\pi_j)^2$	$\frac{(P_j - 100\pi_j)^2}{100\pi_j}$
20-40 years old	158	60.77	42	18.77	352.28	8.39
40 and older	102	39.23	58	-18.77	352.28	6.07
					Total = 14.46	
					$f_j / 100 = 2.6$	
					$\chi^2 = 37.6$	
					Crit χ^2 (p=.05) = 3.84	

In fact, in 1991 Leming asserted that the profession was “graying” (Leming, 1991, 227) owing to the fact that only 11% of the teachers were under the age of 30, and 32% were under the age of 40. Just as the profession was graying, he wrote that teachers were more experienced. In 1986, the median years of teaching experience was 15 with 30% of secondary teachers with 20 or more years of experience (227). Nearly the reverse is true today where the majority of respondents in the current study (65.5%) claimed to have taught less than 12 years, and only 18.4% claimed to have taught 20 years or more.

The demographic trends among respondents to the survey are commensurate with more recent analyses of teaching demographics, which indicate that while a growing proportion of the teaching force is over 50, the teaching force of those under 30 is growing as well (Zumwalt & Craig, 2005). However, in Massachusetts those in the older

ranges of the teaching population appear to have begun to retire from teaching, leaving many novice and newer teachers at the helm in schools. Given that this relatively young teaching force is likely to be a vocal bloc in Massachusetts school districts in the upcoming decade or two, this finding should have significant bearing on policy decisions made for and by this teaching force. This is discussed at greater length in chapter six.

Finally, the Massachusetts HSS teachers studied were significantly more likely to declare themselves as liberal or very liberal than social studies teachers nationwide.

Table 22: *Political Ideology Comparisons between Massachusetts and Nationwide HSS Teachers*

Political Identification	Massachusetts	Nationwide
Liberal	49.8%	15%
Moderate	42.1%	57%
Conservative	7.3%	28%

Note. Nationwide figures were obtained from Leming, J. S. (1991). Teacher characteristics and social studies education. In J. P. Shaver (Ed.), *Handbook of research on social studies teaching and learning* (pp. 661). New York: McMillan Publishing Company.

Leming examined quantitative data on teachers' political orientations and found that social studies teachers tended to be proportionate to both the general teaching force as well as the larger American population (Leming, 1991). Specifically, NCEI data found in the late 1980s that the majority of teachers (57%) identified themselves as moderate and the next largest group (28%) identified as conservative (229). He notes that varying types of response categories (e.g. "liberal," "tend to be liberal," "Democrat," "conservative,"

Republican,") made it difficult to draw general conclusions across different survey instruments. However, he declared, the conservative to liberal ratio generally fluctuated between 1.5 and 2.0 to 1 (Leming, 1991, 229).

Table 23: *Chi-Squared Goodness of Fit Test between Massachusetts and Nationwide HSS Teachers in Terms of Political Ideology*

	Observed		Expected %			
	f_j	P_j	$100\pi_j$	$P_j - 100\pi_j$	$(P_j - 100\pi_j)^2$	$\frac{(P_j - 100\pi_j)^2}{100\pi_j}$
Liberal	132	49.8	15	34.8	1211.04	80.74
Moderate	110	42	57	-15	225	3.95
Conservative	19	7.3	28	-20.7	428.5	15.3
Total = 99.97						
$f_j / 100 = 2.6$						
$\chi^2 = 259.97$						
Crit χ^2 (p=.05) = 5.99						

Again, quite the opposite is true in the current study where the largest group of respondents (49.8%) declared themselves to be very liberal or liberal, and only 7.3% declared themselves to be conservative or very conservative. Part of the stark difference between the data compiled by Leming in the early 1990s and the data gathered in the current study might be a reflection of the times. Specifically, the late 1980s saw many Americans, even moderate and liberal Americans, voting for the very popular Republican

president Ronald Reagan. Regardless, it is commonly understood that Massachusetts is a blue state. Here, the data suggest that Massachusetts teachers are also disproportionately liberal.

In almost all of the characteristics analyzed above, Massachusetts teachers are unlike HSS teachers nationwide. Teachers in this sample are younger, more liberal, and more likely to align with the reflective inquiry orientation to teaching than HSS teachers nationwide. The implications of this are considered in the final chapter.

Massachusetts Teachers Seem to Work from a Theory

Education researchers have made repeated claims that teachers in the field of history and social studies suffer from a lack of consensus about why they teach, what they teach, and how they teach. For example, Stodolsky & Grossman (1995) contended that of the five academic subjects in the K-12 curriculum, social studies is among the most poorly defined. Marker and Mehlinger (1992) argued that HSS has endured a “continuous and rancorous debate” about the purpose of the field (p. 832). Thornton (1991) wrote that without an explicit effort to choose a theory to work from, the HSS curriculum will continue to be threatened by fragmentation and incoherence (p. 242). While the notion may be true that social studies teachers *nationwide* do not work from a uniform theory about the purpose, pedagogy and content of the subject, the same should not be said about teachers in Massachusetts. My data indicate that teachers in Massachusetts are in relative agreement about the purpose, pedagogy and content of social studies, and that their approach is more critically-oriented than teachers nationwide. As I discuss below, this consensus may be a very good thing in that it could

make the task of education policy workers and test developers within the state much easier.

The Nuanced CT Teacher in Massachusetts: More centrist than nationwide HSS teachers

While previous empirical studies of orientations of HSS teachers showed that HSS teachers nationwide tend to be more “critically” oriented than “cultural transmission” oriented (Anderson, et al., 1997; Leming, et al., 2006; Martorella, 2001; Vinson, 1998), this sample of Massachusetts HSS teachers are unique in the overwhelmingly large proportion of critically-oriented teachers. These teachers don’t just endorse RI teaching philosophies and approaches, the vast majority of them actively endorse teaching pupils to become agents of social change. Even the unusual and disproportionately small group of CT oriented teachers (<5%) in this sample are not as conservatively oriented as similarly oriented teachers nationwide.

Most studies similar to this one have found that the CT orientation is by far, the most easily defined and readily observed tradition of HSS dating back to the late 1970’s (Abowitz & Harnish, 2006; Anderson, et al., 1997; Barr, et al., 1977, 1978; Bennett & Spalding, 1992; Goodman & Adler, 1985; Leming, 1991; Martorella, 2001; Morrissett & Haas, 1982; Seixas, 2000). This is not to say that teachers who work from a CT orientation constitute an overwhelming majority of HSS teachers nationwide. The claim that the CT orientation is the most easily defined and readily observed tradition of HSS teaching is in reference to the CT orientation’s archetype status in the HSS teaching field. In other words, while teachers who work within this orientation may only be a sizeable minority, when many people think of history and social studies teachers, they may be

likely to picture the teacher who works in the CT orientation. The items on the survey instrument that were intended to appeal to teachers working within the CT orientation are equivalent to, and in some cases reproductions of items that have been used successfully to identify teachers working within this orientation for over thirty years. However, teachers who work within a CT orientation consistently comprise a smaller percentage of sampled nationwide HSS teachers in these studies than teachers who work from a critical orientation. In fact, in the three quantitative studies of HSS teachers' orientations, teachers working within the CT orientation were found to comprise 23% (Anderson et al., 1997), 20% (Morrissett, 1977), and 9.5% (Vinson, 1998) of the sampled HSS teachers.

This study finds that the CT orientation in Massachusetts was difficult to identify, and comprised an even smaller proportion of the Massachusetts HSS sample (<5%), than studies of nationwide samples of HSS teachers. Very few teachers working from a CT orientation in Massachusetts favorably rated items that were intended to appeal to teachers in their orientation. In other words, the items intended to measure the CT approach (i.e. "Students should believe democracy is the best form of government." "Teachers should assess students primarily with objective tests that parallel course content.") did not create predictable differentiation among this sample. CT oriented teachers had low total scores on all five of the scales indicating that they did not respond favorably to any scale but they disagreed to a lesser extent with items meant to measure the CT orientation.

However, although Massachusetts HSS teachers who were characterized in this study as CT oriented did not respond favorably to CT items, they did respond strongly

and negatively to items intended to appeal to teachers in the RI orientation. In theory, the RI and CT orientations espouse opposing beliefs regarding the role of the student in the classroom and the larger purpose of history and social studies. For example, Massachusetts HSS teachers who were characterized in this study as CT oriented disagreed with RI items on the “Purpose and Pedagogy of HSS” scale that claimed that students should “View American institutions and traditions with a critical eye” (Item 8), and “demonstrate understanding of how to succeed personally & academically” (Item 14) and that teachers should plan lessons that “ask students to practice skills of democratic activism” (Item 22). Massachusetts HSS teachers who were characterized in this study as CT oriented also disagreed with RI items in the “Content Choice” scale that claimed that content should “emphasize the histories of minority groups” (Item 27) or “be student-selected based on students’ own needs and interests” (Item 33). It is not surprising that the RI items, which focused on including students in content decisions or choosing non-canonical content, created clear differentiation in this sample of Massachusetts HSS teachers. Teachers who disagreed with these questions most closely resembled the CT orientation, but they were a modification of the traditional CT orientation.

There is reason to believe then that the CT orientation, as Barr, Barth and Shermis originally proposed it in the 1970s, exists among the Massachusetts HSS teaching force today, but the Massachusetts practitioners who were characterized as ascribing most to this orientation did not fit neatly into the discrete category envisioned by previous scholars. When viewed as points on a field (Figure 3), these teachers, while still in the CT quadrant, are much closer to the center of the field.

The Nuanced RI Orientation in Massachusetts

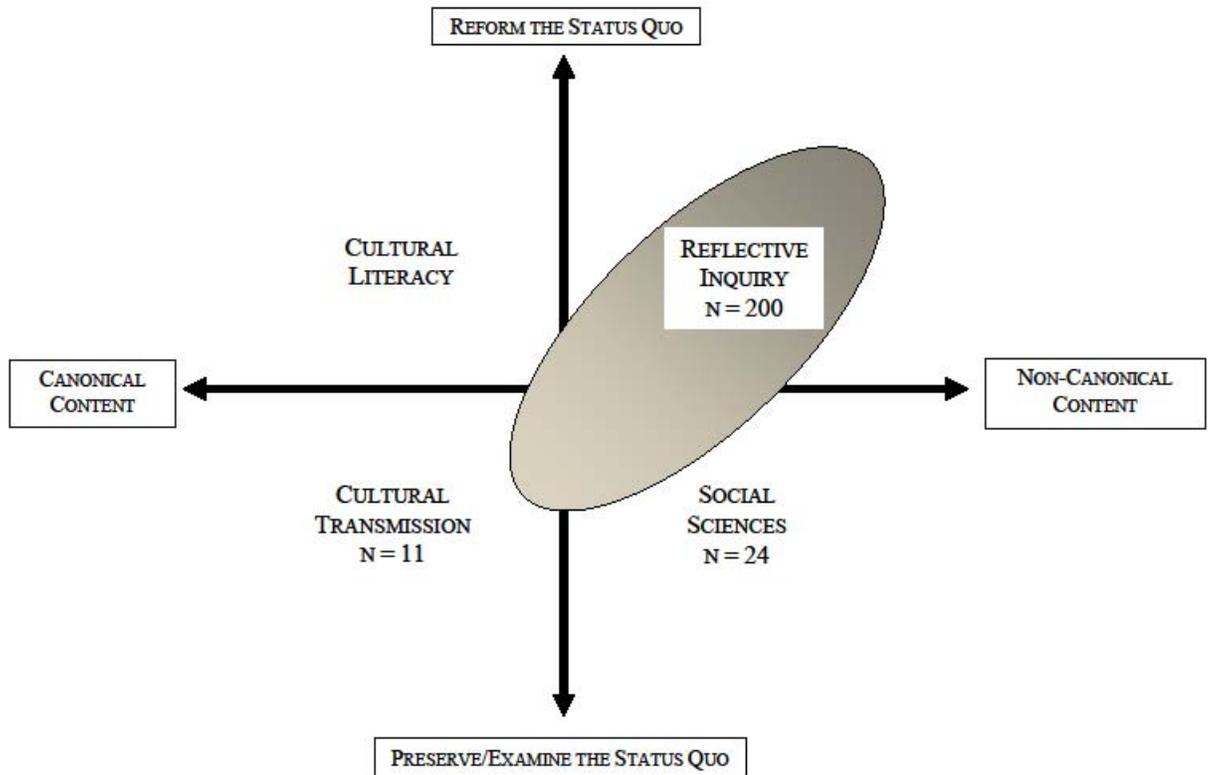
In addition to the CT orientation, teachers in this sample who work within the RI orientation presented an anomaly as well when compared to HSS teachers nationwide surveyed in previous studies. The findings of four earlier quantitative analyses of HSS teachers (Anderson, et al., 1997; Leming, 1992; Morrissett & Haas, 1982; Vinson, 1998) showed that teachers disproportionately identify with critical and reflective inquiry orientations over more conservative cultural transmissions orientations (47%, 52% and 63.1% respectively). However, in the current Massachusetts study, HSS teachers favored the critical RI orientation at a higher frequency than earlier studies (nearly 80%).

Furthermore, the vast majority of teachers working from an RI orientation in this sample agreed with survey items about the reformation of society and the introduction of non-canonical content in the classroom. For example, on the “Purpose and Practice of HSS” scale, these teachers agreed or strongly agreed with statements such as “Students should work toward social justice and equality” (Item 11) and teachers should plan lessons that “allow time for students to discuss power inequity in the United States” (Item 17). These teachers also agreed or strongly agreed with items on the “Preferred HSS Content” scale, including Item 34, that states that content should “emphasize the histories of minority groups” and “focus on past and contemporary examples of racism, sexism and elitism and how to reduce/eliminate them in the future” (Item 34). The finding that Massachusetts HSS teachers were disproportionately more critical and disproportionately less CT-oriented than the national samples previously studied lends further credibility to

the notion that Massachusetts teachers are unlike the national population in their critical leaning.

When these findings are displayed on a field map of orientations, a clear pattern emerges. Figure 3 shows the field map with a large oval spanning the center to the top right quadrant of the teaching orientations. The coloring in the oval symbolically represents the frequencies of Massachusetts teachers aligning with the orientation where the darker shades of gray represent higher frequencies. Massachusetts HSS teachers who most identify with the CT orientation were fewer in concentration (<5%, n=11), and further up and to the right in the quadrant compared to other nationwide teachers working from a CT orientation. Meanwhile, teachers working from an RI orientation in Massachusetts were unlike teachers working from an RI orientation nationwide in that they were the vast majority of the sample (77.7%) and in that the majority of the teachers working from an RI orientation explicitly acknowledged they wished to educate students to be social change agents.

Figure 3: *Field Map of Orientations with Massachusetts HSS Teachers (n =235)*
Overlaid



In many earlier empirical studies of history and social studies teachers’ orientations, education researchers study random samples of HSS teachers generated by the National Council for the Social Studies. This methodological approach has been heavily critiqued by Leming (1992, 2003, 2008) who claims that the professional organization is largely skewed toward liberal (and therefore critically-oriented teachers). To avoid that type of sample error, this study used a sampling technique that randomly sampled from all secondary history and social studies teachers in Massachusetts. The

intention was to obtain a more accurate sample of HSS teachers – one that reflected a larger population of politically conservative and cultural transmission-oriented teachers as Leming argued. Even with this effort to obtain a balanced field of HSS teachers, the result was an overwhelming number of Massachusetts teachers who favored the critical reflective inquiry orientation (nearly 80%). Likewise, teachers in this sample were far more likely to declare themselves to be liberal or very liberal (49.8%) than conservative or very conservative (7.3%).

The Many Similarities between Democratic Pragmatism and Reflective Inquiry

Democratic pragmatism served as the theoretical frame through which I analyzed the results of this study. Here, the theoretical lens serves to ground the first of three comparisons I make in this dissertation. In the first comparison, I explore how the pedagogical and curricular visions of teachers who work within the RI orientation (an orientation overwhelmingly favored by this sample of Massachusetts HSS teachers as indicated on the survey) compare to the ideals of teaching history and social studies outlined and promoted by educational theorists who work from a democratic pragmatist theory. In chapter five, I present the second and third comparisons. For the second, I compare the skills and constructs measured by the proposed Massachusetts test in United States history with the ideal outcomes (skills and constructs) proposed by those who work from a democratic pragmatist theory. In the final comparison, I analyze the relationship between HSS teachers' orientations and state standards and assessments in History and Social Studies.

From the theoretical perspective of democratic pragmatism, finding an overwhelming majority of HSS teachers' who claim to teach students from a particular orientation is a good thing. This type of overwhelming professional consensus about what should be taught, how it should be taught, and why it should be taught, should be used to inform and validate educational policy. This is especially true in light of the fact that, above all other orientations, teachers working from an RI orientation claim to teach the skills those from the democratic pragmatist perspective value, and to the same ends that they envision. Specifically, the skills of democratic deliberation are central to goals of RI oriented teachers.

Democratic pragmatist theory holds clear pedagogical messages for teachers in a democratic society. In fact, Gutmann (1987) maintained that "Democratic education begins not only with children who are to be taught but also with citizens who are to be their teachers..." (p. 49). While theorists from the democratic pragmatist perspective often focus their discussion on the appropriate *outcomes* of democratic education – in other words, what characteristics a graduate must exhibit – one may gather a sense of the type of pedagogical approaches they are likely to endorse. To that end, the question as to what pedagogy best encourages democratic character is addressed.

Most philosophers who describe themselves as pragmatists consider the real effect or practical outcome of any activity as centrally important for divining both meaning and truth. Determining the worth of an educational intervention then can only be assessed by examining its outcome: does the outcome have any practical utility in a democratic society? "Practice, rather than theory," Festenstein (2004) asserts, "is at the heart of

knowledge...Reasoning is understood as a process of deliberation; that is, as a goal-directed activity” (p. 292). The most practical skill a young democrat can learn, according to many democratic pragmatists, is the skill of effective inquiry – that is the process of encountering a problem, developing an effective method to investigate the problem, reasoning through potential solutions and, ultimately, finding resolution to the issue. While the acquisition of concepts and skills of reasoning are both seen as key objectives in education, the skill of reasoning is supreme. Meanwhile, the skill of methodical inquiry is as central to the task of the reflective inquiry teacher as the name of the orientation suggests – inquiry is front and center.

Problems in democratic societies require citizens who are adept at this skill because, as most democratic pragmatists contend, effective deliberation and inquiry – that is the quality of the debate, discussion and persuasion that occurs over a problem – is “crucial for the legitimacy of the outcome” (Festenstein, 2004, 294). Or, as Gutmann and Thompson (1996) put it, the “moral authority of the collective judgment depends in part on the moral quality of the process by which citizens collectively reach those judgments” (p. 4). Teachers in a democratic society must enact pedagogy that serves the purpose of training youth in the skills of inquiry. In a sense then, teachers should start with the end goal in mind and work backward to determine the best pedagogical intervention for achieving that aim.

Alexis deTocquville noted that the American citizens he observed did not acquire democratic skills through formal training and theory alone. Rather, these democrats gained knowledge by *practicing* democratic skills – as one may do by sitting on a jury

(Tocqueville, 2000, 262). Dewey also supported this form of democratic training often arguing that the best way to learn is to do. For Dewey, the argument that one learned best through *doing* had specific implications for history and social education. He claimed, “The only way to make the child conscious of his social heritage is to enable him to perform those fundamental types of activity that make civilization what it is” (Hickman & Alexander, 1998, 232). In this sense, Dewey’s version of effective pedagogy is problem centered, perhaps involving simulations of democratic controversies that would require the student to develop their inquiry skills to resolve the problem (Hickman & Alexander, 1998, 229).

The pedagogical approaches of teachers working from an RI orientation speak directly to these democratic requirements. Teachers working from an RI orientation agreed or strongly agreed that “teachers should allow time for students to discuss power inequality in the United States” (Item 17), that they should “plan lessons that allow students to inquire into their own problems” (Item 18) and “plan lessons that ask students to practice the skills of democratic activism including discussion and debate” (Item 22). The very orientation that the vast majority of Massachusetts HSS teachers claimed to work from, reflective inquiry, represents something of a pedagogical incarnation of the philosophy that democratic pragmatists endorse.

What impact have the nearly 80% of HSS teachers had in Massachusetts? The fact that the vast majority of Massachusetts HSS teachers teach from a particular orientation appears to have little bearing on what policy actors do when initiating or instituting education policy. This is borne out by the fact that, throughout the literature

review process, no studies were uncovered by state-level education policy leaders studying the teaching orientations of HSS teachers within public schools. This begs the question of whether test developers at the Massachusetts Department of Elementary and Secondary Education are testing students for the types of knowledge and skills that HSS teachers claim to be teaching, that they are testing students for the knowledge and skills that theorists who work from a democratic pragmatist philosophical stance endorse, or something else entirely.

CHAPTER FIVE: A STATE US HISTORY TEST WITHOUT A PATH FORWARD

The Massachusetts Comprehensive Assessment Systems' 10th/11th grade test in United States history (MCAS-US test) has had a rocky history. In 1997, the Massachusetts Department of Elementary and Secondary Education (MassDESE) released history and social studies frameworks. In 1999, they piloted a version of the MCAS test that was intended to become fully operational after 2002. This test was comprised predominately of global history items. For various reasons including public outcry and abysmally poor student performance on the pilot tests, test development of the global history MCAS ended in 2002. Test developers then started the redevelopment of the MCAS-US frameworks and test. The newest version of the frameworks was released in August 2003. The standards-based test that was developed focused chiefly on United States history (60%), civics (25%) and economics (15%) and each item was linked explicitly and directly to the standards (*Massachusetts History and Social Science Curriculum Framework*, 2003). For example, the following item appears in the MCAS Guide to History and Social Science Assessments (2007). Below the test item example is the standard from which the item was derived.

Sample Test Item 1: *Item 3 from the MCAS Guide to History and Social Science Assessments (2007)*

*Which of the following is a **legal** requirement of all United States citizens?*

- a. To provide aid to the poor*
- b. To vote in federal elections*
- c. **To serve on a jury if called***
- d. To work for the community*

Curriculum Framework Learning Standard for Item 4

HS.USI.19 Explain the rights and responsibilities of citizenship and describe how a democracy provides opportunities for citizens to participate in the political process through elections, political parties, and interest groups.

The MassDESE piloted this assessment for several years, and each year they claimed that the test would be operational or “high stakes” within the next few academic years. At the outset of this study in 2007, test developers in Massachusetts were ready to administer the last pilot test of the MCAS-US test to the state’s 10th and 11th graders before the test was to become fully operational, or “high-stakes,” in 2011.

However, citing budgetary concerns, the Commissioner of Education in Massachusetts requested in February 2009 that the Board suspend the spring 2009 pilot administration of the MCAS-US and subsequent operational administrations of the test for two years. The Board approved the Commissioner’s request, and state workers are now set to pilot the test again in 2011 (Vaznis, 2009b). Though some mention has been made of revising the test, reports indicate the main objective of postponing the test was

the onset of the budgetary crisis and not some larger construct or philosophical problem with the MCAS-US test (Dwyer, 2009; Vaznis, 2009a, 2009b). For that reason, it seems likely that the test that will become operational in 2011 will not deviate much from the sample items that the MassDESE provides to the public on their website.

This chapter addresses the constructs, concepts and skills measured by the MCAS-US test by analyzing what Porter (2008) refers to as the “content message” of the test. In the first section I discuss how the analysis was done, including a description of sample items from the MCAS test that were rated and analyzed, as well as a description of the rating process and the members of the MCAS-US test rating panel. I next present the findings from the content analysis of the MCAS-US test. In the third section, I discuss what the content message of the MCAS-US test means for its efficacy as an educational tool. Also, I make the second of three major comparisons in this study between the content message of the test and the educational tenets of democratic pragmatism, a perspective consistent with increasing civic knowledge and participation. In the final section of the chapter, I make the last of three major comparisons in this dissertation between Massachusetts HSS teachers’ orientations toward the subject, on the one hand, and the skills and constructs measured by the MCAS-US test, on the other. Here, I compare the findings from the survey of Massachusetts HSS teachers with the findings from the content analysis of the MCAS-US test.

Process for Content Analysis

Obviously, content analysis of any test requires access to test items. Though the MassDESE planned to make pilot tests available to the public the summer after each

administration, to date, test items have largely been kept confidential. Test developers at the state level argue that test items that are both clearly linked to state standards and that generate psychometrically sound differentiation within the test population are difficult and expensive to devise (Dwyer, 2008a). In fact, the MCAS Guide to History and Social Science Assessments (2007) states that items take approximately two years to develop because they “undergo extensive review and field-testing” (p. 6). The MCAS guide explains how sample items are tested rigorously, aligned stringently with state standards, and reviewed by expert panels called “Assessment Development Committees” (ADCs). These assessment panels are comprised of Massachusetts classroom educators, curriculum developers and school administrators (Dwyer, 2008a; *MCAS Guide to History and Social Science Assessments*, 2007, 8). The ADCs review each item to determine whether or not the item is aligned with a standard, concept or skill from the Massachusetts Curriculum Frameworks and whether it is appropriate to grade level (Dwyer, 2008a; *MCAS Guide to History and Social Science Assessments*, 2007, 8).

Given limited availability of items, the content analysis for this study was based on sample items that the MassDESE released to the public via two online sources. First, MassDESE created a guide to the MCAS-US Test (*MCAS Guide to History and Social Science Assessments*, 2007) that has sample items from each of the three social science and history tests (grade 5, 7, and 10/11). A second set of test items was made available from the MassDESE’s website (*High School U.S. History Released Items*, 2007). Both sources provided descriptions of the proportions of test item types (common items, module items, open-response items) and point values for each.

What I call “the sample test,” which was used in this study for the content analysis, refers to the compilation of all of the 10/11th grade items from those two sources. The sample test was compiled, distributed to, and rated by, each of the content analysts who were recruited for this study to analyze items on the sample test (Appendix F). Of the 36 items on the sample test, three were open-ended response questions (i.e. eliciting an extended written response from students) and were worth four points each. The 33 remaining items were multiple choice questions and were worth one point each. Items on the sample test were worth a total of 45 points.

The sample test given to the content analysts for this study differed in a few ways from the actual pilot test, which was the version of the test administered to 10th and 11th graders over the past several years. While the pilot tests that were administered to students were never made public, an evaluation can be done of the similarities and differences between the sample test devised for this study and the pilot. One way that the sample test differed from the pilot test was that the MassDESE claimed that eighty percent of the pilot test was multiple choice questions and twenty percent was open ended questions. The pilot test contained both “common items” and “matrix sampled” items for both the multiple choice and the open-ended questions. Common items were items that appeared on all pupils’ tests across all districts. Matrix sampled items were experimental test items and were different on each districts’ test throughout the state. When the MassDESE did the pilot tests, only common items were scored. The sample test contains only common items and has a slightly lower proportion of multiple choice items and a slightly higher proportion of open ended response questions. An additional difference

between the sample test compiled for this study and the pilot tests administered to 10th and 11th grade students (but not available publicly) was that the pilot high school test was comprised of 44 common items and four open-ended items for a total of 64 points. What this means is that the sample test booklets comprised for this study, including the 33 multiple choice and three open ended response questions, represented about 75% of the full-length pilot tests that students have taken in the past.

As described in chapter three, I drew on the work of Porter and colleagues (Porter, 2006; Porter & Polikoff, 2008; Porter & Smithson, 2001) to conduct the content analysis. Porter and colleagues have used content analyses to assess standards and high stakes test items based on the depth of thinking required of students to answer questions and the content areas that the items tap. To determine the depth of thinking required by students, Porter and colleagues assign an item or standard to one of five levels of “cognitive demand.” These cognitive demand categories rank the type of performance tasks that students must carry out to demonstrate their knowledge from “memorizing” or “solving routine problems” on the basic end of the spectrum to “solving non-routine problems,” and “conjecturing,” or “proving” on the more complex end of the spectrum (Porter, 2006). Porter and colleagues also assess the standards and test items in terms of what specific content area they tap within a subject. To analyze the content that an item or standard taps in social studies, for example, they devised a complex list of general content topics common to social studies – a list with 31 separate categories such as “Principles of American Democracy” and “American Constitutionalism” and over 325 subcategories such as “majority rule vs. minority rights” and “landmark Supreme Court

cases (e.g. *Marbury v. Madison*, *Brown v. Board*, and *Miranda v. Arizona*)” (Wisconsin Center for Education Research, 2008).

The team then used this information to verify that the items on the test proportionally reflect the content and skills that are required of students according to the states’ standards or curriculum frameworks. In other words, Porter and colleagues use their method to analyze alignment between state standards and state assessments. Though Porter and colleagues’ previous work was used as a starting point for the creation of content analysis methods for this study, two major adjustments were made to their procedure and are outlined below.

One adjustment to Porter and colleagues’ approach included changes made to how depth of knowledge was assessed. As stated, in studies conducted by Porter and colleagues, categories of “cognitive demand” were used to rank items or standards. However, in this study, I used four “Depth of Knowledge” (DOK) levels (Figure 4) derived from the work of Norman Webb at the University of Wisconsin, Madison (Webb, 2002), which were preferable to Porter’s cognitive demand categories for two important reasons. First, Webb acknowledged the similarity between the DOK levels he developed and the cognitive demand levels used by Porter (Webb, 2009). However, he claimed DOK levels are more useful for content analysis while the cognitive demand categories provide a framework for analyses of mental operations (Webb, 2009). In other words, the DOK levels can be used to analyze items while the cognitive demand categories require the analysis of individuals’ cognitive engagement on a task. Secondly, MCAS-US test developers at the state department of education stated informally to me that the

MassDESE used Webb’s taxonomy (Figure 4) to rate items on all of the MCAS tests in terms of the DOK they require (Dwyer, 2008b). Webb’s taxonomy was chosen, therefore, in an effort to maintain consistency between how the test developers within the MassDESE evaluate the MCAS-US test and this study.

Figure 4: *Four Depth of Knowledge Levels Derived from Webb*

Literal	→			Deep
Level One: “Recollection”	Level Two: “Basic Reasoning”	Level Three: “Complex Reasoning”	Level Four: “Extended Reasoning”	
Recall, recognize, identify, list, define, and/or reproduce information. Describe and explain information in maps, charts, tables or drawings	Basic reasoning, contrast and compare, convert, classify or sort. Descriptions & explanations result in understanding how & why.	Complex reasoning, justify how & why. Apply using evidence. Use concepts to solve problems. Propose & evaluate solutions.	Extended reasoning. Plan, investigate or develop something over time. Connect & relate ideas & concepts within & among content areas.	

For my study, a second adjustment was made to Porter and colleagues’ approach. This complex list of content topics, which included a list of over 300 subtopics, was replaced for this study with five categories reflecting the five teaching orientations to HSS (CT, SS, RI, RIPE, RISE).

Figure 5: *Five Content Area Categories Reflecting the Five Teaching Orientations*

Cultural Transmission	Social Sciences	Reflective Inquiry	Reflective Inquiry for Personal Ends	Reflective Inquiry for Social Ends
Focuses on discrete factual and/or canonical knowledge about history & government including American heroes, symbols & dates.	Focuses on content and problems in the social sciences (i.e. economics, sociology, psychology, sociology, geography).	Focuses on social problems, controversial issues, critique of authority, broad range of political content.	Emphasizes building relationships between people. Developing self reflection, empathy and human relations	Emphasizes minority/multicultural groups, racism, sexism, power, classism, inequality, and how to bring about social justice and equality.

This change was made to ease the task of the content analysts and to allow for comparisons between content found on the test and content favored by teachers in each of the teaching orientations previously discussed. It should be noted that when the survey analysis was done, it was determined that these five categories should be collapsed into three categories, and that was subsequently done for the content analysis as well. Specifically, the categories RI, RIPE and RISE were subsumed under one category, reflective inquiry.

Though Porter and colleagues' process has been used in many different ways by several groups to analyze standards and assessments, Porter described one particular content analysis approach in detail. In studies conducted by the American Institute for Research (AIR), three raters independently assigned each specific part of a content standard to both a cognitive demand, and a content topic cell within a matrix (see Appendix E) (Porter, 2006). Data gathered from individual raters were then averaged together cell by cell and proportions were created for each cell by dividing by the sum of the average weights across all cells in the matrix (Porter, 2006, 143). Procedures vary across studies, as does Porter's approach, depending on if the researchers are analyzing standards or teachers' practice.

Following the procedures outlined by Porter, four content analysts were recruited purposively for this study from a graduate school of education, including one tenured professor and three advanced doctoral candidates in the field of history and social studies education. Each content analyst had previously worked with secondary history and social studies students and teachers, and all were familiar with the HSS content typically covered at the secondary level. The content analysts were given the packet of sample MCAS-US test items (Appendix F) and a blank content analysis matrix (Appendix E) which showed the four categories of Depth of Knowledge (DOK) along the x-axis, and five content area categories along the y-axis. Panelists were then given oral instructions about how to rate each of the items along with a description of each of the rating categories on the x and y axis of the grid. The first sample item from the MCAS-US test was completed as a group with discussion. This activity was intended to generate

clarifying questions and establish a common understanding among the raters about how to proceed. Raters were then asked to assign each of the 36 sample MCAS-US items to one of the four DOK categories, and one of the five content area categories. Again, while they were given a content analysis grid with five content area categories along the y-axis, these five categories were subsequently collapsed to three categories once the survey analysis was done.

Frequencies of items generated across the four completed content alignment grids were then loaded into SPSS where raters were treated as cases and ratings for content type and depth of knowledge for each item on the MCAS-US test were treated as variables. To assess the degree of inter-rater reliability between and among the four raters, an Intraclass Correlation (ICC) coefficient was generated, which is appropriate for measuring inter-rater reliability with two or more raters. The test generates the same .0 to 1 Cronbach's alpha coefficient range and that coefficient is interpreted much the same way as a Cohen's Kappa coefficient. In this two-way mixed model, I looked for a measure of consistency across the average of all ratings assigned by members of the panel. The test generated a Cronbach's α coefficient of .878. The four member panel rated the 36 items with a high degree of consistency on both dimensions.

Some questions arose while panelists were rating sample items in the content alignment grid. Decisions were made about how to deal with these issues based on what the four panelists believed Porter and colleagues and those who used the content analysis method at the MassDESE would have decided in a similar situation. The panelists reasoned that this would facilitate the comparison of findings between my study and

content analyses previously conducted at the MassDESE. First, panel members had a difficult time differentiating between Level One DOK, “Recollection,” and Level Two DOK, “Basic Reasoning.” Many items on the MCAS-US test required the recall of very specific historical facts, which were then to be used to reason out a solution to a given problem. One panel member noted that if the student taking the test were not able to recall the specific historical fact, he or she would be “foiled from the start.” The question was whether this kind of item should be considered a “recall” item or a Level One DOK item. Again, in an effort to ease comparisons between the content analysis completed of the MCAS-US test by the MassDESE and my study, panelists agreed that they would try to employ the rating rationale likely used by MassDESE raters. Even though we all agreed that a student would be “foiled” if he or she could not recall particular information, we reasoned that test developers at the state level who rated each item’s DOK had likely assigned items to the highest DOK category the item tapped. Panelists thus decided to rate the items based on the highest DOK category the item tapped.

A second issue arose regarding the content categories along the y-axis. Panel members were confused about whether or not certain items addressing multicultural content should be categorized as RI or CT content. For example, one panelist argued that although one item dealt with the African American experience in America (i.e. African American’s role in World War Two or desegregation efforts at the turn of the 20th century), these events have become an accepted part of the American historical canon. Should those items then be considered items favored by a cultural transmissionist? A discussion ensued, and it was pointed out that the fact that this panelist would consider

the African American experience part of traditional American history might be a reflection of her particular stance toward teaching history even though this has not come to be accepted as a central part of the American historical canon by most HSS teachers. Theoretically, for many cultural transmissionists, social history – including multicultural history – is not a content topic that is highly endorsed. For that reason, panelists were asked to rate any item about multicultural history or social history as “RISE” content because it was content that would likely be endorsed by teachers who use reflective inquiry for social ends. The RISE orientation was eventually collapsed (along with the RIPE orientation) into the reflective inquiry (RI) orientation.

A third issue arose as to how best to deal with open-ended response items. There were three open-response items in the collection of sample items. Each open-response item had two parts. For example, for one item (Sample Test Item 2), students were provided a copy of Lincoln’s *Gettysburg Address* with the following prompts:

Sample Test Item 2: *Item 27 from the High School US History 2007 Pilot Test*

Released Items

In the Gettysburg Address, President Lincoln refers to the nation as having “a new birth of freedom.”

- a. Explain what President Lincoln meant when he called for “a new birth of freedom.” You may support your answer with information from the Gettysburg Address and your knowledge of American history.*
- b. Explain why Lincoln believed a unified nation should be preserved. You may support your answer with information from the Gettysburg Address and your knowledge of American history.*

The two different parts of this item appear to tap two different levels of DOK. For that reason, open-ended questions with two parts were treated as two items with the first above called 27a and the second 27b. Additionally, because open ended response items were given four times the weight that a single multiple choice item was given when proportions of test items were determined, each of these two items was worth two points each. The affect of weighting the open response items is discussed further in this chapter.

Finally, it must be noted that Webb’s Level Four DOK, “Extended Reasoning,” is described as requiring students to work on a problem over a length of time. Assessment tasks that could tap the Level Four DOK category include things like portfolio work or other long-term projects. For that reason, no item on the MCAS-US test was able to reach Level Four DOK on Webb’s taxonomy. This calls into question the rationale for choosing such a framework. However, a Massachusetts state-level test developer confirmed the

inability of any item on the MCAS-US test to tap the highest fourth level of Depth of Knowledge (Dwyer, 2008b). Despite this, state test developers persisted with this framework.

Findings from Content Analysis

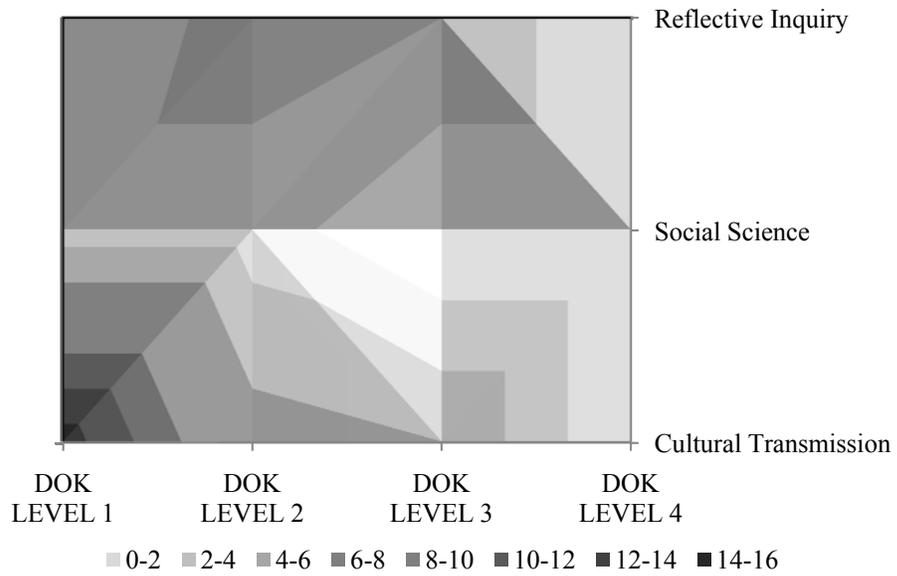
As mentioned previously, each rater was given a content analysis grid and asked to rate items on two dimensions. Following Porter and Polikoff (2008), data from these grids were used to create surface area charts to provide a visual representation of the 36 items according to the two dimensions. Figure 6 shows the frequency distribution of items for each of the four raters separately. The surface maps that were generated read like a topographical map showing peaks and valleys of frequencies of items rated in each category. As is the case with Porter's work (2006), in each of the four surface maps below, "shading represents relative content emphasis and is analogous to altitude on a topographical map" (p. 143) with darker shades being higher concentrations of emphasis and lighter shades indicating less concentrations of emphasis. Unlike Porter's previous work, these surface maps show far fewer gradations of gray; the shapes are much more angular. This is due to the small number of items ($n=36$) rated by the four panelists in this study compared to the hundreds of items rated in Porter's analyses. Additionally, as described earlier, there were far fewer categories for the two dimensions compared to Porter. The differences between my method and Porter's work had very little impact on the interpretation of the surface maps for my study because, though a little less refined, it was still possible to make out the topographical nature of the surface maps. In the surface

maps generated for this study, the legend outlines the density of gray and the corresponding range of item points associated with each gradation of gray.

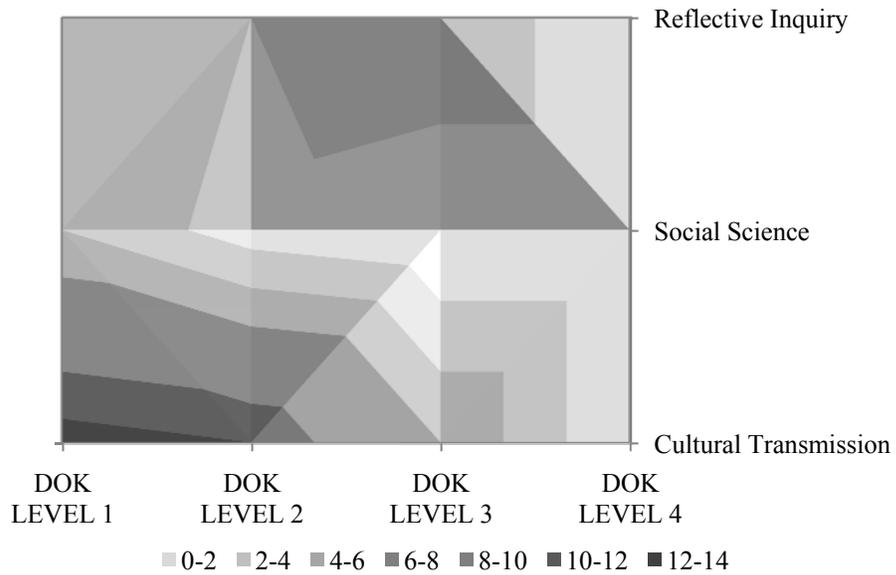
What is first made clear by these surface area maps (Figure 6) is the high degree of inter-rater reliability among Raters One, Two, Three and Four across the 36 test items, even though Rater Four was somewhat of an outlier. Each of the four surface maps, including the outlier, shows a mountain peak in the bottom, left hand square as indicated by the darkest gray colors. In other words, all members of the panel rated the highest frequency of points on the MCAS-US test in the cultural transmission content category with a Level One, “Recollection” or Level Two, “Basic Reasoning” Depth of Knowledge. The surface maps for Raters One, Two, and Three also showed a concentration of medium gray in the top, middle-right corner. This represented a high frequency of points on the sample MCAS-US test tied to reflective inquiry items that were rated as Level Three, Complex Reasoning DOK. Rater Four is a bit different because while other raters tended to rate most items as cultural transmission, this rater viewed those same items as social science or reflective inquiry content because of the item’s economic or social message. Regardless, this rater rated the biggest proportion of items on the sample test as cultural transmission in content focus similar to the other three raters.

Figure 6: *Frequency Distribution of Points across the Two Dimensions by Rater*

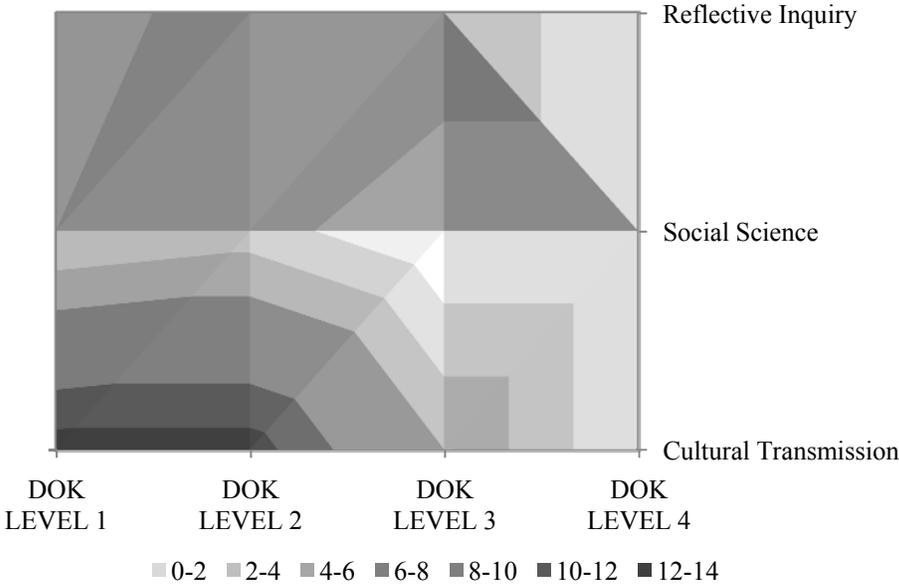
Rater One



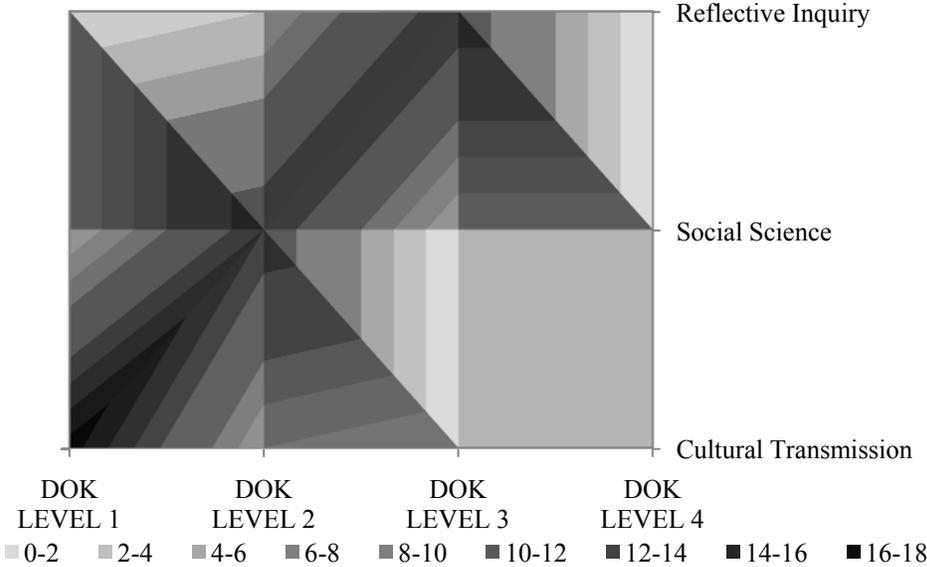
Rater Two



Rater Three



Rater Four

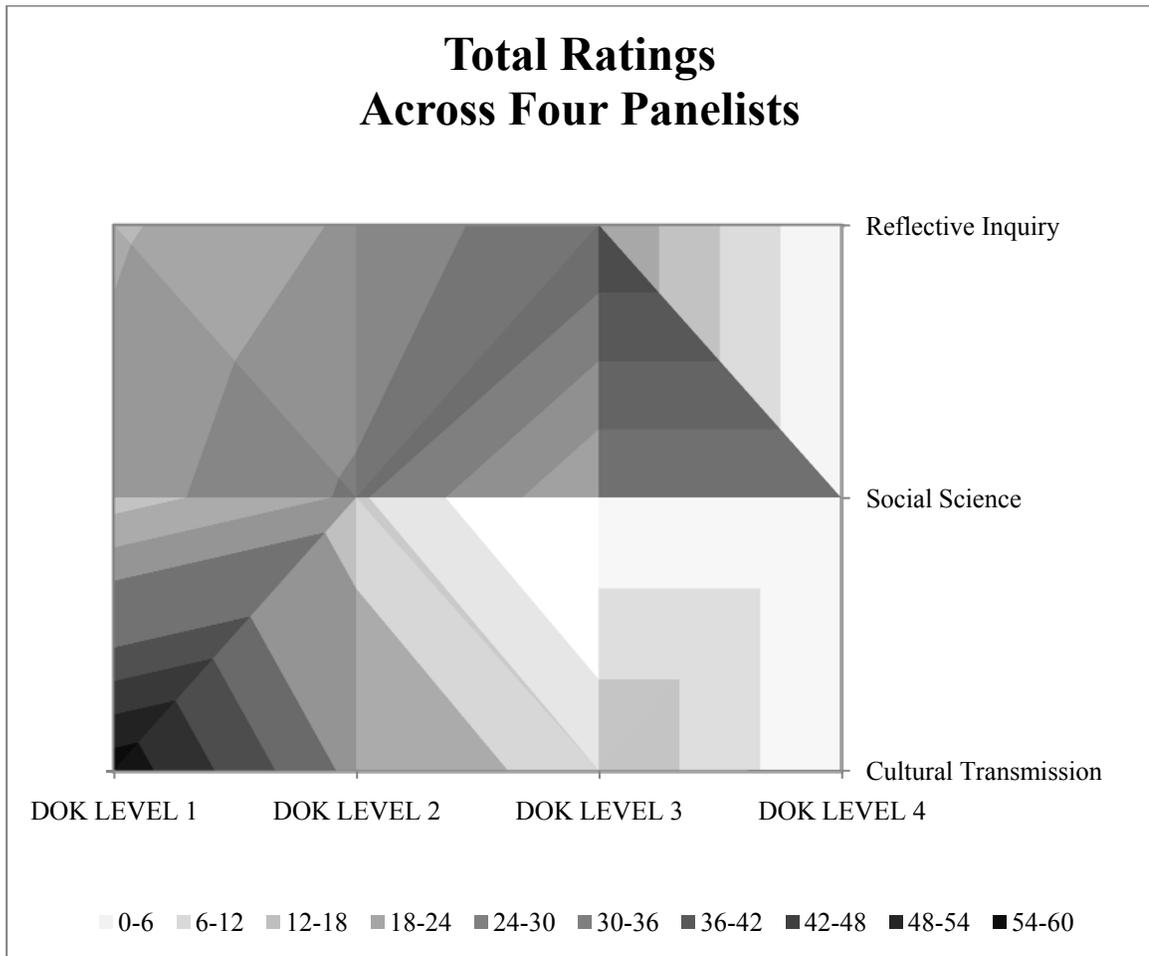


What these surface maps visually demonstrate is that the highest frequency of points on the sample MCAS-US test (the darkest gray/black) are cultural transmission or social science content that tap only the first or second DOK level. In other words, the items on the sample MCAS-US test are assessing a limited range of content and depth of knowledge.

Ratings for each of the four panel members in each content and DOK category were then combined into total categories to generate one overall surface map (Figure 7). The same coloring was used for point values in that the darkest gray areas have the highest frequency of points while the lightest gray areas have the lowest frequency of points. Figure 7 shows in even sharper relief the four raters' tendency to rate the highest proportion of items as cultural transmission content with a DOK Level One, "Recollection," or Two, "Basic Reasoning." This is represented by the mountain peak in the bottom left hand corner of each of the surface maps. A second, smaller peak is located at the top, middle of the surface area. This represents the smaller, but high number of point values associated with reflective inquiry content items that were rated as DOK Level Three, "Complex Reasoning." Were item frequencies across the two dimensions to be examined without giving weight to the open ended response questions – that is to say, if all of the 36 items were weighted equally at one-point each – these distributions would look markedly different with even greater concentration of items in the lower left quadrant. Or, put another way, the items which boosted point values in the RI category were open ended questions. If the open ended questions were given equal weight to the multiple choice questions, the RI category would have much shorter columns meaning

that the RI items on the test would be an even smaller proportion of the test than they were in this sample test. These items are referred to as “fat” or “heavy” items, and the significance of this finding is discussed further in the following section.

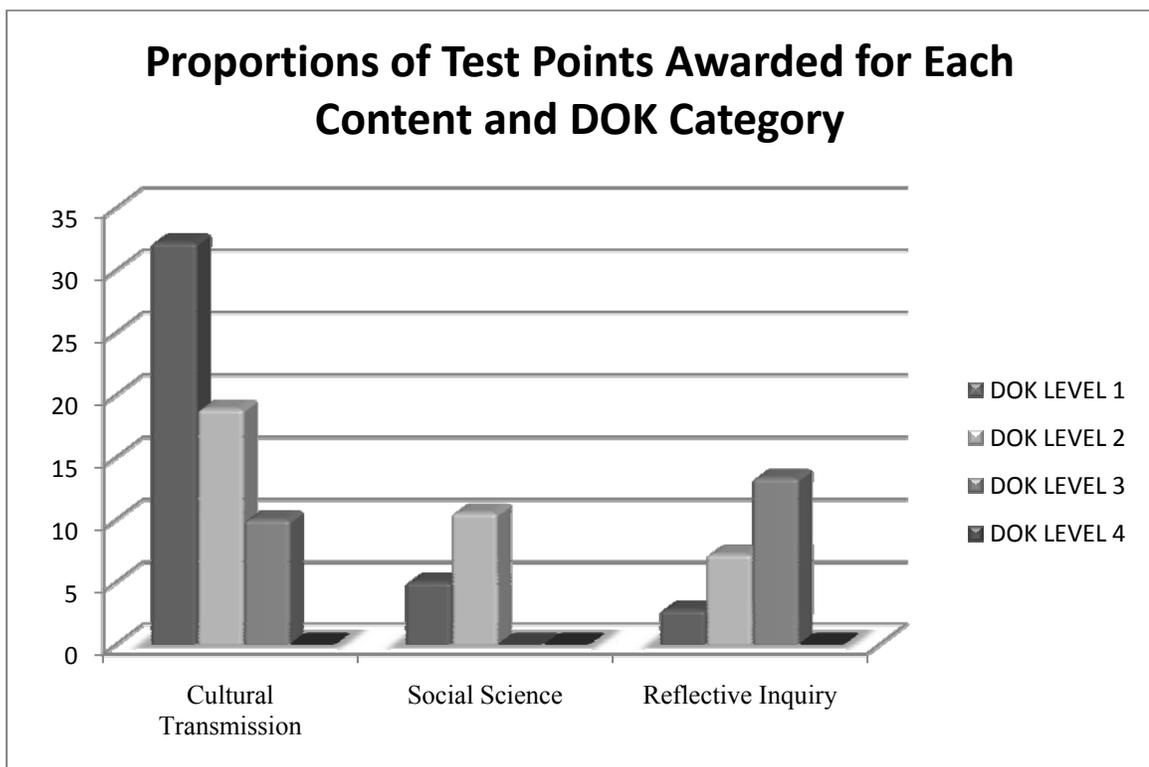
Figure 7: *Total Frequency Distribution of Points across All Four Panelists*



Each of the above figures shows frequencies in terms of how many *points* are being rewarded by each of the questions. In other words, there are 36 sample items that were rated. Thirty-three items were multiple choice items and therefore worth one point each. Three of the 36 items were open-ended items with two probes. Because each of the open ended questions was worth four points, each of the two probe questions was worth

two points. The 36 sample items were worth 45 points total. When the scores of the four raters for each of the 36 items are combined, the 45 point test was worth 180 points. Proportionally, of those 180 total points, 61.1% was content most associated with a cultural transmission orientation to HSS, 15.6% was content most associated with a social science orientation (specifically they were economics questions), and 23.4% was content most associated with reflective inquiry (Figure 8 and Figure 10). Cultural transmission content questions, therefore, accounted for more than half of the sample test items.

Figure 8: *Proportions of Test Points Awarded for Each Content and DOK Category*

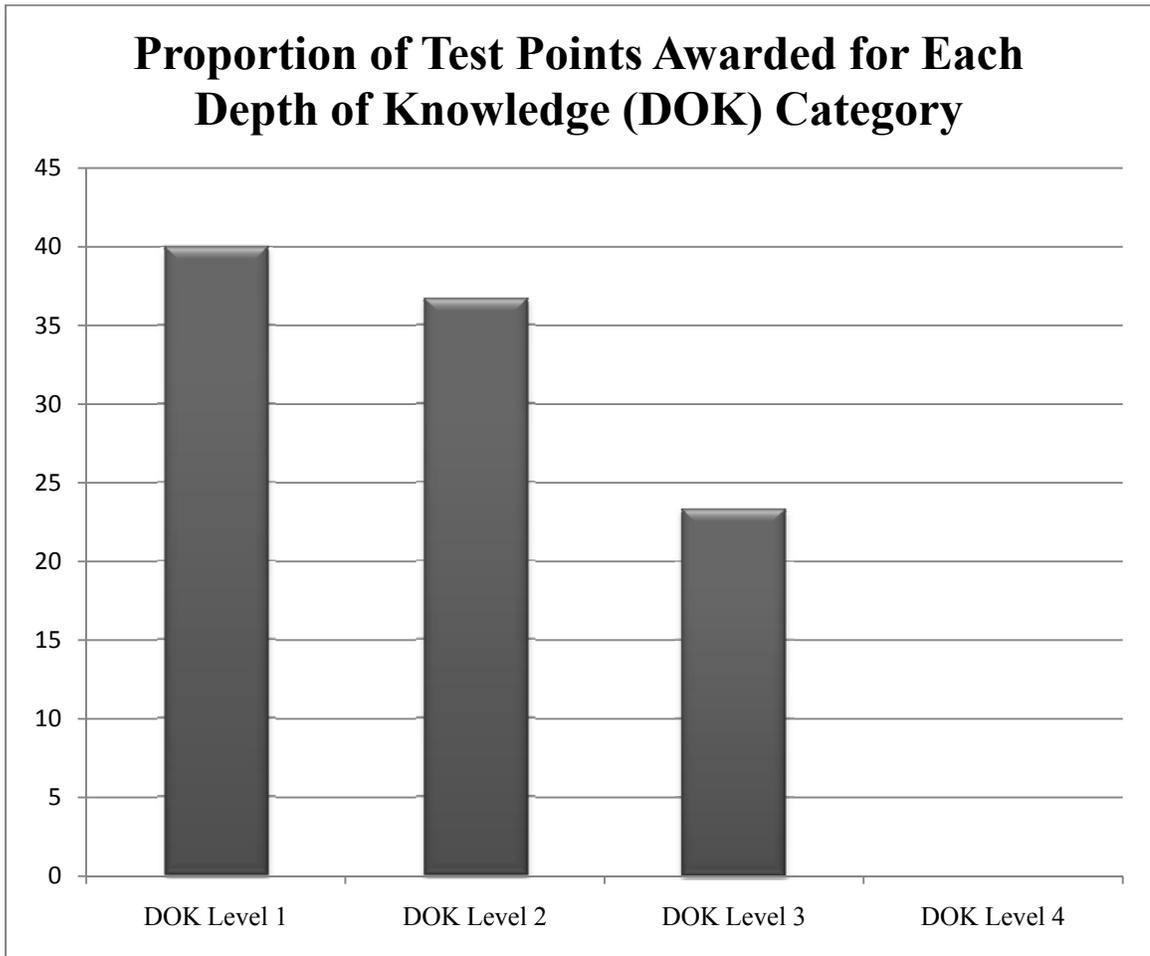


In terms of Depth of Knowledge categories (Figure 8 and Figure 9), 40% of the total points were associated with items rated as Level One or “Recollection,” 36.7% of the total points were affiliated with items rated as Level Two or “Basic Reasoning,”

23.4% of the total points were linked to items rated as Level Three or “Complex Reasoning,” and as mentioned previously, no points on this test could be rated as Level Four because no items on this test are capable of tapping what Webb would describe as Level Four, “Extended Reasoning.”

Figure 9: *Proportion of Test Points Awarded for Each Depth of Knowledge (DOK)*

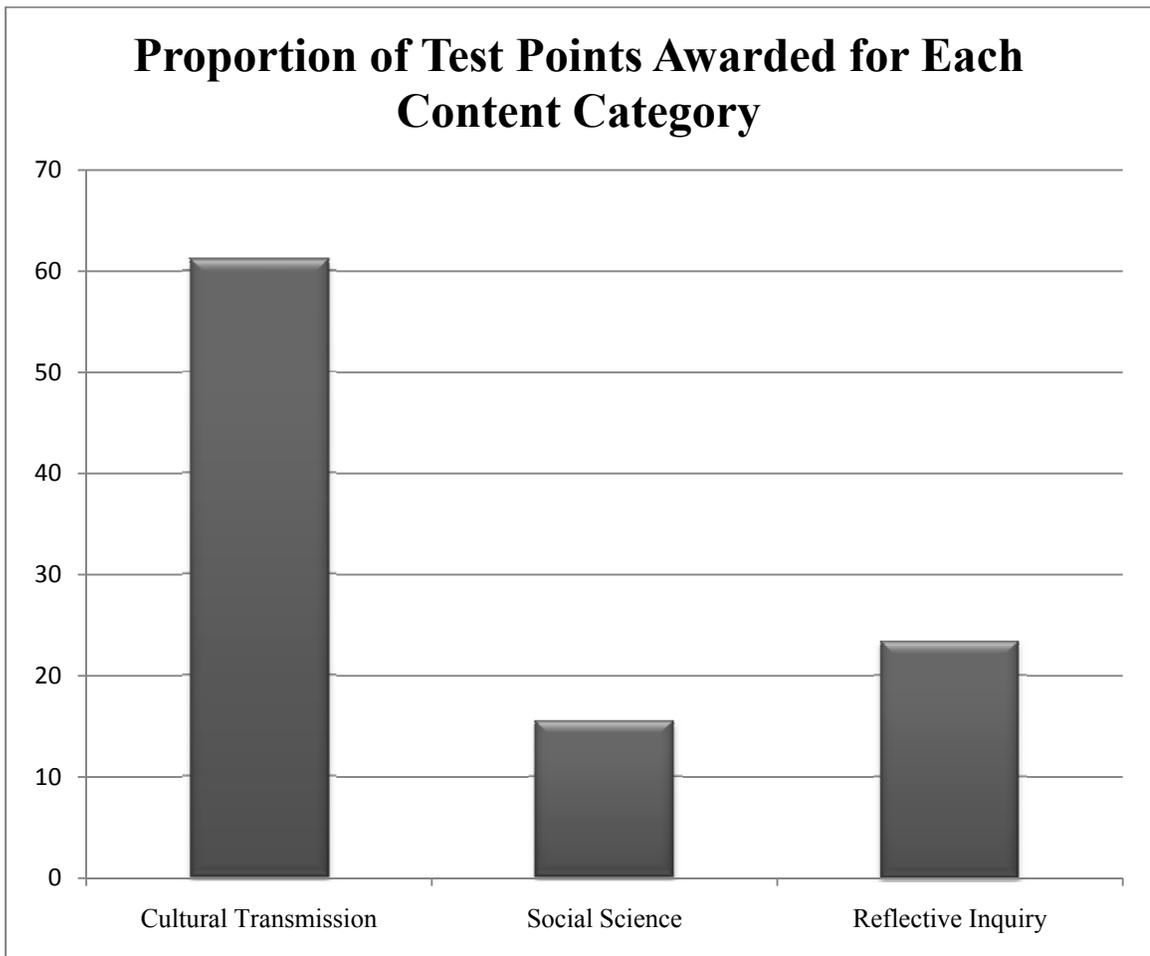
Category



As demonstrated by the surface maps as well as the percentage of points awarded, the majority of items on the MCAS-US test were items that were associated most with

cultural transmission content (61.1%), and a Level One, “Recollection” DOK (40%). In other words, these items focused on discrete factual, traditional and/or canonical knowledge about history and government and included important American heroes, symbols and/or dates. Also, most items required only that students recall, recognize, identify, list, define or reproduce information.

Figure 10: *Proportion of Test Points Awarded for Each Content Category*



To illustrate the type of content knowledge and depth of thinking required by items rated as cultural transmission content and Level One DOK, below I include three items (Sample Test Items 3, 4 and 5) that all panelists rated in those categories:

Sample Test Item 3: *Item 1 from the MCAS Guide to History and Social Science Assessments (2007)*

In his Farewell Address, President Washington warned the United States to avoid which of the following?

- a. Trade relationships with European nations*
- b. Diplomatic relations with European powers*
- c. **Permanent alliances with European powers***
- d. Immigration agreements with European nations*

Sample Test Item 4: *Item 11 from the MCAS Guide to History and Social Science Assessments (2007)*

*Who among the following **most** influenced the passage of the Bill of Rights?*

- a. John Adams*
- b. Benedict Arnold*
- c. **James Madison***
- d. Roger Williams*

Sample Test Item 5: *Item 22 from the High School US History 2007 Pilot Test*

Released Items

The Supreme Court case Marbury v. Madison (1803) established which constitutional principle?

- a. Due process*
- b. **Judicial review***
- c. Implied powers*
- d. Equal protection*

Among all of the categories across the x and y axes in this content analysis, panelists were most often in agreement about items that fell into the CT content and Level One, “Recollection” DOK categories. Panelists had less agreement about items that tested social science or reflective inquiry content, and were less consistent when distinguishing between Level Two, “Basic Reasoning” and Level Three, “Complex Reasoning” DOK items. Yet, even while they were less consistent, several raters still made similar deductions about these items such that inter-rater reliability coefficient (as stated previously) was .878. For example, three out of four panelists rated the following item (Sample Test Item 6) as reflective inquiry content, and Level One, “Recollection” DOK:

Sample Test Item 6: *Item 8 from the MCAS Guide to History and Social Science Assessments (2007)*

Important Events in the United States Women's Rights Movement

The timeline shows the following events:

- 1963: *The Feminine Mystique* is published
- 1964: Civil Rights Act is passed
- 1966: National Organization for Women is founded
- 1971: Equal Rights Amendment is submitted to Congress
- 1972: Title IX of the Higher Education Act is passed
- 1973: *Roe v. Wade* is decided

Who wrote The Feminine Mystique, a book that helped spark the women's rights movement?

- a. **Betty Friedan**
- b. Phyllis Schlafly
- c. Eleanor Roosevelt
- d. Sandra Day O'Connor

In other words, three out of four of the panelists believed the content tested in this question emphasized minority and/or multicultural groups, thus making it content favored by a Reflective Inquirer. However, the item only required students to recall, recognize, identify, list, define or reproduce information. Unless students had memorized the fact that Betty Friedan had authored the famous manuscript, they could not engage higher depths of knowledge to reason the correct answer, they could only guess.

Panelists rated nearly 38% of the items as Level Two DOK, Basic Reasoning. However, as stated previously, there was less agreement about which items fell into this DOK category. For example, three out of four panelists rated the following item (Sample Test Item 7) as Level Two DOK:

Sample Test Item 7: *Item 16 from the High School US History 2007 Pilot Test Released Items*

The excerpt below is from Frederick Douglass’s “Independence Day” speech in 1852.

The rich inheritance of justice, liberty, prosperity and independence, bequeathed by your fathers is shared by you, not by me. The sunlight that brought life and healing to you has brought stripes and death to me. This Fourth of July is yours, not mine.

Frederick Douglass,

“Independence Day” speech (1852)

In this speech, Frederick Douglass was speaking on behalf of which group?

- a. Exiled American Indians*
- b. Deported American Jews*
- c. Oppressed Irish Americans*
- d. Enslaved African Americans*

Most panelists rated this item as Level Two DOK, “Basic Reasoning,” meaning that the item required students to use basic reasoning to determine the correct answer to the question. In other words, students might not be familiar with Frederick Douglass or know of his “Independence Day” speech, but they could reason that he was speaking on behalf of enslaved African Americans based on the content of the excerpt.

Often when panelists were in agreement about the items that fell into this DOK category, they showed less agreement about which content orientation the item tested. That is to say that while they would rate the item Level Two DOK, the four panelists were split as to whether the item tested cultural transmission, social science or reflective inquiry content. It seemed as though the more complex DOK the item reached, the harder it was for the panelists to uniformly identify what content the item was assessing. This may show that items that required deeper levels of knowledge also required a broader range and type of content knowledge. For example, all four panelists rated the following item (Sample Test Item 8) as Level Two DOK but two of the panelists rated it as social science content and the other two rated it as cultural transmission content:

Sample Test Item 8: *Item 2 from the MCAS Guide to History and Social Science Assessments (2007)*

How did the cotton gin affect the Southern economy from 1800 to 1860?

- a. It encouraged industrialization in the South.*
- b. It promoted economic equality in the South*
- c. **It strengthened Southerners’ reliance on slavery.***
- d. It increased Southerners use of indentured servants.*

Content analysis indicated that the majority of points on the sample MCAS-US test came from items that were cultural transmissionist in content orientation and Level One, “Recollection” DOK. The next largest proportion of points on the MCAS-US test also came from items that were cultural transmissionist in content orientation and Level Two, “Basic Reasoning” DOK. A much smaller proportion of points on the sample test were from items designated social science or reflective inquiry in content orientation or Level Three, “Complex Reasoning” Depth of Knowledge. Most of the point value here was a result of the heavy nature of the question (i.e. they were open ended response questions and therefore worth four, rather than one point each), *not* because there were multiple items in these categories.

Content analysis of the sample test items raises questions about whether the MCAS-US test is, in fact, measuring the constructs, concepts and skills that are consistent with the goal of increasing civic knowledge and participation. In the next section, I make the second of three comparisons of this dissertation. Specifically, I compare the findings from the content analysis with the goals outlined by educators who work from a democratic pragmatist framework. I suggest that the prevalence of multiple choice items on the test, coupled with the predominance of traditional, “consensus” historical content is misaligned with the goals of history and social studies education that promotes greater civic engagement.

Discussion of the Content Analysis Findings

It is commonly assumed that standardized tests are scientific, factual and objective. Yet, while test development experts are becoming increasingly better at the

task of assessing pupils' academic ability (Dwyer, 2008a), the products of their efforts are never perfect. Tests are not created in a social vacuum. Often, there are subtle messages in tests about how subject matter is prioritized by the many stakeholders involved in test development. As stated previously, Porter used the term "content message" to describe the underlying theoretical construct that links individual items on a test together. These content messages, not just the psychometric properties of individual items, are part of the quality of a test as a whole.

Content analysis of the sample MCAS-US test provides a clear picture of the content message of the piloted MCAS-US test: The test has a cultural transmissionist content message, which has been critiqued by those who work from a democratic pragmatist perspective. In the following sections, I compare the content message of the MCAS-US test against this educational vision in order to address the problem of diminishing social capital and democratic knowledge. I suggest that, given the prevalence of multiple choice items and the test's focus on traditional historical content knowledge, the MCAS-US may work in opposition to the goal of increasing civic knowledge and democratic participation.

Multiple Choice Items in the MCAS-US Test

Standardized test developers' overreliance on one type of psychometric measure, the multiple choice question, has troubled education researchers for decades. Several previous researchers claim that standardized history tests are often comprised solely of multiple choice items (Kurfman, 1991; Pahl, 2003a, 2003c; Rothstein, 2004; Wineburg, 2004). This is not a neutral observation. Rather, these studies claim that the prevalence of

multiple choice items represents a disconnect between the goal of teaching social and historical research and thinking skills and the test content.

The 36 items on the sample MCAS-US test were all multiple choice save three open-ended response questions. In other words, 33 of the 45 total points (73.3%) on this sample test were from multiple choice items while only 12 of the 45 total points (26.6%) were from open-ended responses. This high proportion of multiple choice items is to be expected given that the MassDESE maintains that all pilot tests are 80% multiple choice and 20% open ended questions. MCAS-US test developers have repeatedly claimed that the test item types found in the sample MCAS-US test are proportionately identical to the test item types found in actual pilot administrations of the full test (*High School U.S. History Released Items, 2007; MCAS Guide to History and Social Science Assessments, 2007*). Therefore, there is reason to conclude that the reliance on this question format was deliberate, and future administrations of the MCAS-US test will similarly assess students predominately with multiple choice items.

While over 73% of the points on the MCAS-US sample test were multiple choice questions, this study found that many of the items (36.7%) pushed students beyond fact recall and required students to demonstrate basic reasoning skills. This contradicts previous studies that equate multiple choice items with fact recall (Cohen, 2008; Kurfman, 1991; Pahl, 2003a, 2003b, 2003c; Rothstein, 2004; Volger, 2003) It might be encouraging for some to know that these multiple choice items tapped the Level Two, “Basic Reasoning” DOK. Their optimism would be short lived, however. Unless students have memorized the endless array of facts spanning hundreds of years of United States

history and dozens of economic concepts that are included in the state frameworks, it is unlikely that students would have the background knowledge needed to demonstrate their basic reasoning skills on the MCAS-US test. For example, the following item (Sample Test Item 9) was rated by all raters on the panel as a Level Two, “Basic Reasoning” DOK.

Sample Test Item 9: *Item 1 from the MCAS Guide to History and Social Science Assessments (2007)*

The quotation below is from a speech given by Senator Albert Beveridge in 1898.

Hawaii is ours; [Puerto] Rico is to be ours; at the prayer of her people Cuba finally will be ours; in the islands of the East... the flag of a liberal government is to float over the Philippines... The Opposition tells us that we ought not to govern a people without their consent. I answer the rule... that all just government derives its authority from the consent of the governed, applies only to those who are capable of self-government.

Senator Albert Beveridge,

“The March of the Flag” speech (1898)

Which policy was Senator Beveridge advocating in this statement?

- a. Containment*
- b. Disarmament*
- c. **Imperialism***
- d. Isolationism*

In the sense that this question asks students to reason, based on the evidence provided, what type of policy the Senator advocated, this is a DOK Level Two, “Basic Reasoning” question. However, the student would also need to *recall* the fact that the name given to this policy was “Imperialism.” The answers provided are not descriptions of the policy

(i.e. “Americans expanded the nations’ territory beyond the continental United States.”)

The answers provided are specific labels to refer to foreign policy agendas, and the student must have mastery of both specific facts and basic reasoning skills to answer the question correctly. While the question forces students to higher levels of thinking, at the most basic level this question is a multiple choice question requiring fact retention and recall. Results of the content analysis, therefore, reinforce the conclusions of earlier conceptual and empirical studies on state standards and standardized assessments in history and social studies. Rather than assessing students’ historical thinking and research skills, the MCAS-US test sample items mostly assess students’ ability to recall factual data about economic principles and United States history; further evidence of the tests’ cultural transmissionist content message.

As stated previously, among those who are critical of the overreliance of standardized tests on multiple choice questions are those who work from the perspective of democratic pragmatism. In order to assess the effectiveness of teachers’ pedagogical interventions, theorists who work from a democratic pragmatist perspective argue that assessments in education must get at whether the student has acquired the skills and capacities to participate in society; to inquire, to reason and to deliberate. This assertion has very tangible implications for testing in history and social studies. Revisionist historians, who argue that historical truth is a social construction, call into question the notion of teaching history as a process of presenting facts for students to memorize (Smith, 2006). Other democratic theorists question the notion that citizenship education should have an outcome that is measurable on traditional tests (Popkewitz & St. Maurice,

1991, 29). Dewey's early writings confirm this line of reasoning (Horn, 2006). He stated simply, "Examinations are of use only so far as they test the child's fitness for social life" (Hickman & Alexander, 1998, 231). Any assessment of democratic education should be held to this standard. This raises questions as to whether the dominance of multiple choice items on the MCAS-US test are useful to a student as he or she moves into the social world. While members of a democratic society are often asked to reason, discuss and deliberate over significant social problems, and are often required to demonstrate the verbal, written and other communication skills necessary to convey their thoughts, one could argue they are rarely if ever in need of discrete historical facts that multiple choice items assess.

The American Grand Narrative According to MCAS-US Test

American historians refer to the "grand narrative" as the story history tells about who Americans are as a people. Contests over the grand narrative have centered on whether that narrative should represent a consensus history – meaning history that is generally shared by dominant groups in America, and which moves in a logical progression of advancement, or a more critical narrative – one that shows the cracks, fissures and conflicts in American history that have created obstacles to the advancement of society (Nash, et al., 2000). Previous studies of standards and standardized tests in history and social studies have concluded that tests highlight consensus or traditional history rather than critical, reformist, social, multicultural, local or world history (Banks, 2004; Cornbleth, 2006; Evans & Pang, 1995; Foster & Morris, 1991; Henry, 1987; Kelly, et al., 2007; Schneider, 1995; Sleeter & Grant, 1987; Stern, 1994). Authors of previous

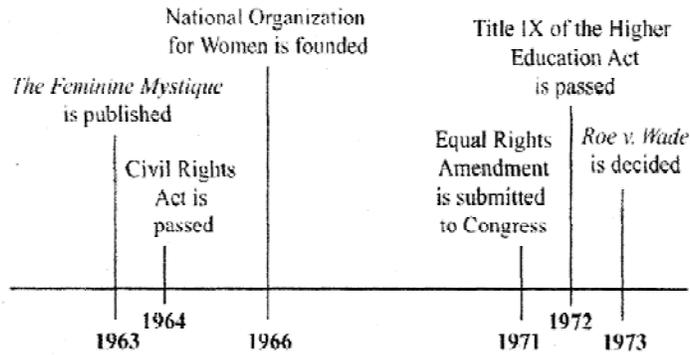
studies argue that when assessments focus disproportionately on traditional history, other voices in the American narrative are ignored since HSS teachers tend to emphasize content that is tested.

The vast majority of items in the sample of MCAS-US test were items that focused on traditional American history including government and military history or economic principles (>75%). Moreover, two of the three open ended response items tested traditional history. Only a very small number of items on the MCAS-US test assessed content that could be considered multicultural or gendered. Even these items required that students recall biographical data about a particular individual (i.e. that Frederick Douglass was an advocate for the rights of African Americans) rather than asking students to think critically about how competing narratives have shaped conceptions about where America has been and what America could become.

While the vast majority of points on the MCAS-US sample test came from items rated as cultural transmission and reinforced the consensus grand narrative, the points that came from items rated as RI content or higher levels of DOK came from a very small number of “heavy” items on the sample test. In other words, the findings of this study are slightly misleading in that the respectable number of points associated with RI content and Level Three, “Complex Reasoning” DOK come from only a few questions on the test. For example, one item (Sample Test Item 10) on this sample test was an open-ended response item to do with the role of women in American history. The open ended question began with a timeline and a description of the timeline and then had two specific questions:

Sample Test Item 10: *Item 10 from the MCAS Guide to History and Social Science Assessments (2007)*

Important Events in the United States Women's Rights Movement



In the 1960s, the women's rights movement regained some of the vigor and enthusiasm that it had lost after the passage of the Nineteenth Amendment, which gave women the right to vote in 1920.

- a. *Identify **one** issue that was important to the women's rights movement in the 1960s and 1970s. Explain why the issue you identified became important to women at this time. Support your answer with information from the timeline and your knowledge of U.S. History.*

- b. *Explain what gains, if any, the women's rights movement has made since the 1960s in resolving the issue you identified in part (a). Support your answer with information from the timeline and your knowledge of U.S. history.*

All panelists rated the content on both parts of this question as RI content. The first question required students' to show their basic reasoning about the issues central to the women's rights movement. This item was rated as a Level Two, "Basic Reasoning" DOK by all raters. The second question asked students to think critically about how the role of women has evolved in their own lifetime. This item was rated as Level Three, "Complex Reasoning" DOK by all raters.

This open-ended question including the two prompts constituted only a very small number of items that had to do with non-traditional American history rated as a high-level DOK item on the sample MCAS-US test. As stated above, the open-ended response items were weighted at four times the point value of the multiple choice items. This single item (Sample Test Item 10) would be similar to what Porter and Polikoff (2008) call a "fat item," or an item that "covers more than one specific [content] topic by cognitive demand combination" (p. 4). In their research, were the fat item worth, for example, three points, but fell into two different content by cognitive demand cells, the point value for that item would be split in half and one and a half points would be placed in each cell the item fell into. In my study, I used a related but slightly different idea. What I call "heavy" items were items that touched on rich non-traditional historical content and required students to show greater depths of knowledge, but simultaneously had a higher point value associated with the item due to the fact that the item was an open-ended response question. This means that one item therefore boosted significantly the frequency of points shown in each of the panelists' RI category. Sample Test Item 10 is a good example of a heavy item because this one question was worth proportionally

more points than other test items, and the item tested both RI content at a Level Three, “Complex Reasoning” DOK.

This finding is problematic given that, since this is a sample test, there is no guarantee that future open-ended response questions will be “heavy” with multicultural or gendered content. In other words, unless test administrators deliberately include items such as this one in upcoming administrations of the test, the point values awarded to the RI category by the panelists may be artificially inflated. Even assuming that some heavy items such as this one will be included in future administrations of the MCAS-US test, still the vast majority of content tested by this sample MCAS-US test was content favored most by the cultural transmissionist orientation – the orientation most aligned with teaching consensus or traditional history.

Critics of tests that portray a traditional version of the American grand narrative contend that there should be balance between “conservative” and “liberal” content. Believing that knowledge is acquired through individual inquiry and shared through a process of democratic deliberation, critics encourage teaching concepts which, over time, have been habitually used to address and solve practical problems. These critics endorse teaching concepts such as “judicial precedence” in order for youth to understand how the Supreme Court and lower courts make complicated decisions about social, ethical, moral and constitutional issues. Or they might argue that in order for students to understand issues such as state power versus individual freedom, they should learn about fundamental documents such as the Bill of Rights and the first three articles of the Constitution. Thus, for these critics, traditional content that reinforces a consensus grand

narrative, including the traditional content outlined in the examples above, has a place in the content required for youth.

The measure of the worth of the item, however, is not simply that these are important concepts that everyone should know, but that they are historical concepts that have *relevance* in contemporary society. In other words, traditional content is not all given equal priority. Only the traditional content that has relevance in that time period is something worth knowing and assessing. The prioritization of these traditional concepts (i.e. what concepts or items are tested and what concepts or items are returned to the vault) is subject to change over time. In a sense then, traditional content is akin to temporally-dependent truths. That is, the relevance of the content and the priority that content is given depend upon current events. Educational theorists Popkewitz and St. Maurice (1991) call this justification of truth “social epistemology.” They place social epistemology in the philosophical tradition of pragmatism because, as with pragmatism, social epistemology acknowledges there are “various ways of knowing [which] are overlapping and continually reconstructed through interactions” (p. 27). Were these critiques of content on standardized tests heeded – including the critiques from those who work from a democratic pragmatist lens – they would have serious implications for how test developers determine what content appears on the standardized test in any given year the test is administered. This is addressed further in the final chapter.

For those who work from the perspective of democratic pragmatism, “knowledge” consists of those concepts that are necessary for easing communication and deliberation between citizens. Common cultural concepts, (e.g. sources of individual

rights such as the Bill of Rights) serves as a starting point from which citizens redefine and reinvent our country's historical narrative (Hickman & Alexander, 1998). While many of the concepts history and social studies standards outline will probably continue to be taught and tested in schools, those who work from a democratic pragmatist lens are likely to argue that the reason for this is *not* that they are established historical truths. Rather, these concepts have been re-justified by the current generation as inter-subjectively chosen historical concepts. With the explosion of new national narratives from historians over the last several decades, it seems appropriate that Americans should prepare themselves for a changing and evolving national narrative that engulfs multicultural, social, critical and/or reformist narratives. From the perspective of democratic pragmatism, this is merely the process of democracy: consensus and revision in historical narratives are to be expected. This suggests that the consensus version of the American historical grand narrative as presented by the MCAS-US test is incomplete at best. Even more conservative neo-pragmatists, who like pragmatists argue that knowledge has to have some practical purpose (Robinson & Groves, 2005), but unlike pragmatists endorse the notion of teaching traditional content in the lower grades, would likely question the lack of critically oriented questions on an assessment meant for adolescents (Rorty, 1999).

Analysis of the Relationship between HSS Teachers' Orientations and the Concepts and Skills Measured by the MCAS-US Test

Many assumptions are commonly made about how standardized tests impact classroom teachers. In chapter two, I used Madaus' (1988) seven principles describing

the perverse and negative impact of standardized tests on the curriculum as a frame to review dozens of articles particular to history and social studies curriculum that make similar claims about the impact of standardized tests in history. However, very few assumptions are made about whether, or how, teachers' curricular orientations are considered when test developers devise standardized tests. Do test developers who create standardized assessments consult classroom teachers to ensure that the tests align with what teachers intended or claim to teach in their classroom? Much more likely is the assumption that test developers create items that are aligned with the frameworks, presuming that teachers faithfully and thoroughly follow each strand of the state frameworks as well.

My analysis of my survey data indicated that HSS teachers in Massachusetts work primarily within the reflective inquiry (RI) orientation. Meanwhile, the content analysis of the MCAS-US test showed that the content message of the MCAS-US test is largely Cultural Transmissionist. In this section, I explore the relationship between the predominantly-RI curricular orientations of HSS teachers in Massachusetts and the predominately-CT content message and constructs measured by the Massachusetts state-mandated U.S. history assessment using data gathered by both the survey of Massachusetts HSS teachers as well as the content analysis of the MCAS-US test as a starting point for this comparison. To address the question, "Is the Content on the MCAS-US Consistent with Teachers' Orientations?" I explore the relationship between teachers' orientations and the content of the MCAS-US test using quantitative data gathered in the content analysis. To address the second question, "How do teachers view state standards

and the MCAS-US test?" I used quantitative data gathered in the systematic survey of Massachusetts HSS teachers.

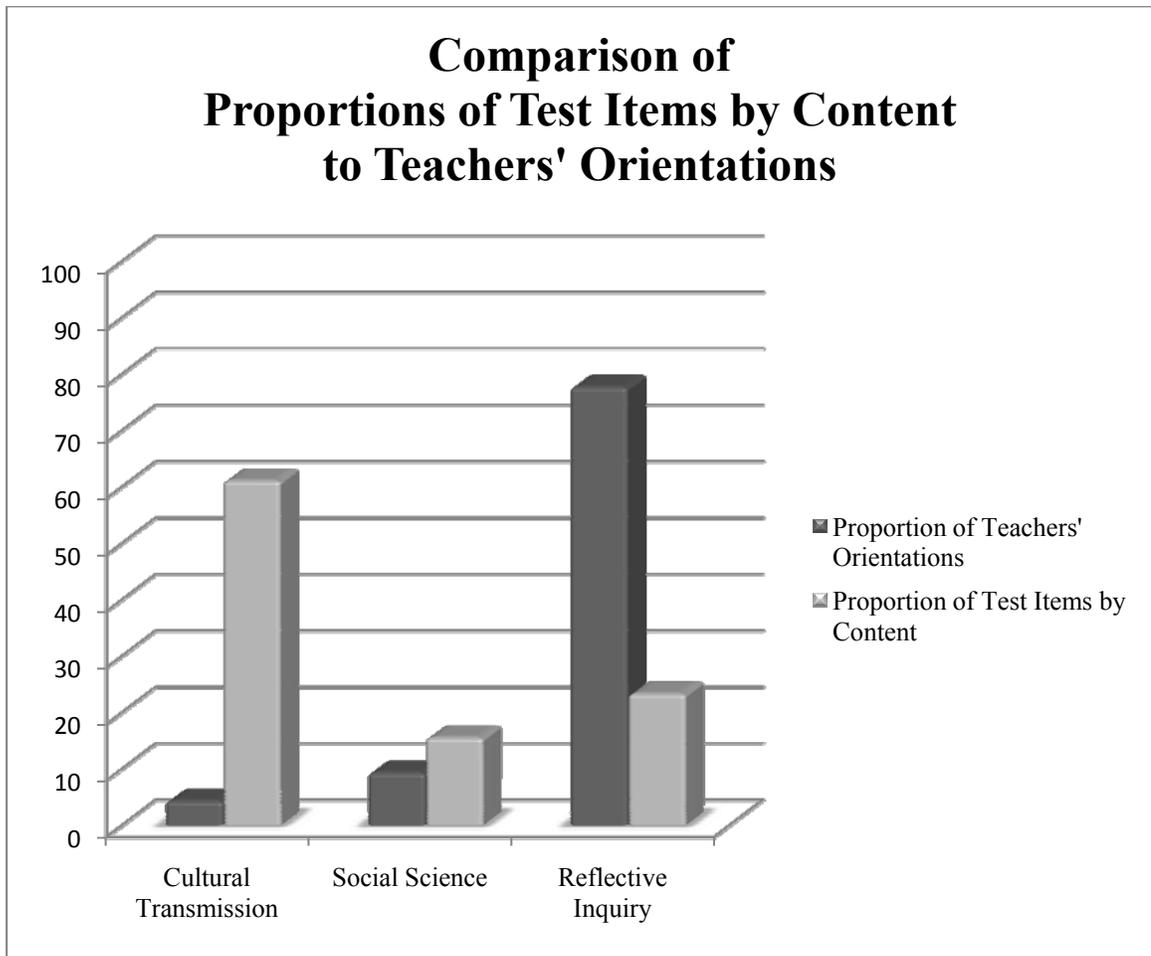
Is the Content on the MCAS-US Test Consistent with Teachers Orientations?

Since the advent of standardized tests, education researchers have been working diligently to determine the "content validity" of those tests. Content validity is a matter of determining if "the content that the instrument contains is an adequate sample of the domain of content it is supposed to represent" (Fraenkel & Wallen, 2006, 153). Some education researchers have studied the content validity of standardized tests by systematically measuring the content message of the tests and comparing that to similar measurements of state frameworks for a particular subject matter. Porter (2006) referred to this as a check of alignment between the "intended curriculum" and the "assessed curriculum" and noted that several reviews of different approaches to measuring alignment have been published (Ananda, 2003; Bhola, Impara, & Buckendahl, 2003; Council of Chief State School Officers, 2002; Olson, 2003; Rothman, 2004). However, Porter also acknowledged that alignment checks can and should be done between intended or assessed curriculum and *instruction* – or what he called the "enacted curriculum." After all, what teachers choose to present in the classroom has a direct bearing on students' opportunity to learn and their subsequent achievement on assessments. My study endeavors to answer a key question: Is the content on the MCAS-US Test consistent with the curricular orientations of classroom teachers?

Using data generated by both the first and second research designs of this study, I compared the curricular orientations of classroom teachers in Massachusetts to the

content message of the MCAS-US test. As discussed earlier, I found that most teachers in Massachusetts align with the reflective inquiry orientation. A small minority of teachers align with either the social sciences (SS) orientation or the cultural transmission (CT) orientation. This heavy leaning toward the RI orientation was true whether or not the sample included teachers whose highest score occurred in two or more orientations indicating that the group of teachers working within the RI orientation was even larger once ties between two RI orientations were accounted for. In fact, the biggest group that had ties between two or more orientations were the three orientations of the RI family (i.e. ties between RI/RIPE, RI/RISE, RISE/RIPE). Meanwhile, I have also shown a very different picture of the proportions of questions on the MCAS-US test. When using the same orientations framework to analyze a sample of questions from the MCAS-US test, I found that the vast majority of the questions were rated as CT and SS items. A very small percentage of the items on the MCAS-US test tapped content most clearly associated with the family of RI orientations. Figure 11 compares the content orientation of items on the MCAS-US to teachers' orientations.

Figure 11: *Comparison of Test Items' Content to Massachusetts HSS Teachers' Orientations*



What is clear is that the MCAS-US test has a very different curricular message than HSS teachers claim to endorse in the classroom. In fact, the figure shows nearly an inverse relationship between content measured on the test and curricular orientations favored by Massachusetts HSS teachers. Over 61% of items on the MCAS-US test were CT content, but only 4.3% of Massachusetts HSS teachers in this sample aligned most with that orientation. Conversely, less than a quarter of the items on the MCAS-US test were RI content, yet nearly 80% of Massachusetts HSS teachers in this sample aligned

with the reflective inquiry orientation. So, while individual test items on the MCAS-US test may be consistent with the Massachusetts curricular frameworks in the sense that the content covered by the item does appear in a more general form in the Massachusetts frameworks, the same cannot be said about content of the MCAS-US test in relation to HSS teachers' curricular orientations. The test does not align with Massachusetts HSS teachers curricular orientations.

I used a Chi-Squared goodness of fit test to determine whether there was a statistically significant relationship between the orientation of items on the MCAS-US test and HSS teachers' curricular orientations. Porter and colleagues (2006, 2008) explicitly use tests of proportionality to determine the alignment between enacted (taught) curricula and assessed (tested) curricula. Here, the null hypothesis is that there is no significant difference between the observed proportions of HSS teachers in each of the three orientations and the expected proportions of items on the MCAS-US test in each of the three orientations. My alternative hypothesis is that there is a significant difference between the observed proportions of HSS teachers in each of the three orientations and the expected proportions of items on the MCAS-US test in each of the three orientations.

$$H_o: \pi_{CT} = \pi_{SS} = \pi_{RI}$$

$$H_a: \pi_{CT} \neq \pi_{SS}$$

$$H_a: \pi_{SS} \neq \pi_{RI}$$

$$H_a: \pi_{CT} \neq \pi_{RI}$$

Table 24 presents the calculations and solutions for the Chi-Squared analysis. In this calculation, Massachusetts HSS teachers' calculated curricular orientations are the

observed frequencies (f_j) and proportions (P_j) and test item orientations on the MCAS-US test as determined by the MCAS-US test rating panel are the expected proportions.

Table 24: *Chi-Squared Goodness of Fit Test between HSS Teachers and MCAS-US Test Items*

	Observed		Expected %		$(P_j - 100\pi_j)^2$	$\frac{(P_j - 100\pi_j)^2}{100\pi_j}$
	f_j	P_j	$100\pi_j$	$P_j - 100\pi_j$		
Cultural Transmission	11	4.68	61.11	-56.42	3184.25	52.10
Social Science	24	10.21	15.55	-5.34	28.48	1.83
Reflective Inquiry	200	85.11	23.33	61.78	3816.32	163.52
					Total = 217.52	
					$f_j / 100 = 2.35$	
					$\chi^2 = 511.17$	
					Crit χ^2 (p=.05) = 5.99	

The Chi-Squared value for this test is 511.17. The critical Chi-Squared value is 5.99 (p = .05). I reject the null hypothesis at the p = .05 level that proportions are equal.

Specifically, the proportions of HSS teachers in each of the three orientations are not

equal to the proportions of items on the MCAS-US test in each of the three orientations. Stated another way, there is a very clear misalignment between the curricular orientations of HSS teachers and the content orientations of sample items on the MCAS-US test on the other.

Given this misalignment, one could come to two possible conclusions. One conclusion is that there is a problem with the content validity of the MCAS-US test. In other words, the MCAS-US test doesn't assess what teachers claim to teach and therefore the test is inadequate. This is supported by the fact that in a thorough review of the literature, no studies were found indicating an effort on the part of test developers to investigate the curriculum that teachers enact in the classroom. This raises questions about whether the test actually assesses what students have been taught. This conclusion is also supported by the fact that MassDESE test developers didn't consider teachers' input about the MCAS-US test until after the test was constructed. MassDESE formed Assessment Development Committees (ADC's) as a way to include teachers in the test development process. Judging by the name, these committees seem to be opportunities for teachers to help *develop* the MCAS tests. However, the development of the assessment was already largely underway by the time teachers were encouraged to participate. Having already formatted the MCAS-US test with predominately multiple choice and a few open ended response items, MassDESE invited teachers to verify that the existing items for the MCAS-US test aligned with the Massachusetts curriculum frameworks (Porter, 2009).^{xi} Teachers did not "develop" the assessment; they helped to verify that the items on the multiple choice test were in line with the standards.

A second and different conclusion that one could reach based on the misalignment of test content and teachers' curricular orientations is that teachers are not adequately and faithfully following the state frameworks. According to MassDESE, the HSS test is constructed based on the Massachusetts History and Social Science Curriculum Framework (2003) (*High School U.S. History Released Items*, 2007; *MCAS Guide to History and Social Science Assessments*, 2007). Test makers, therefore, may not believe they are expected to know what teachers claim to do in the classroom; they only expect that teachers' content and instructional choices are similarly guided by the Massachusetts Curriculum Frameworks. Thus, the test does not necessarily have to be consistent with what teachers do in the classroom because it is assumed that teachers' content and instructional choices are also guided by the Massachusetts Curriculum Frameworks. The assumption that both test makers and teachers use the same standards to guide their work leads to the conclusion that the test and teachers' orientations should be consistent. This conclusion means that if there is misalignment, classroom teachers are not faithfully following the state standards.

However, a third explanation is possible. It may be that both teachers and test makers are faithfully working from the state frameworks, yet the products of their efforts are widely divergent. That is to say, it is possible that both test makers who have created an assessment whose content reflects a cultural transmission orientation, and classroom teachers with a reflective inquiry orientation could both be aligned with the state standards because they have extrapolated differing meanings from the standards in their respective work.

How could this happen? Professional historians and educational researchers agree that the state standards are often so inclusive that no teacher could ever cover the mandated curriculum in 180 instructional days. In fact, in a nationwide study of history and social studies state frameworks (Gagnon, 2003), Massachusetts' standards failed to meet two key criteria. Most notably, the frameworks failed to outline "topics teachable within the allotted timeframe" (Gagnon, 2003, pp., 75). This means that even if teachers were teaching the standards directly and with fidelity, they would be unlikely to move through all of that content within one academic year. They invariably would have to omit certain concepts and content. Meanwhile, the Massachusetts curriculum frameworks are so exhaustive, it would prove difficult for test developers to design items *not* rooted in the frameworks (Gagnon, 2003). In other words, any item written about any topic in United States history could be justified by a standard in the frameworks.

For example, Massachusetts learning standard USII.27 requires students to "Analyze the causes and course of the women's rights movement in the 1960s and 1970s" and lists Betty Friedan and Gloria Steinem as contributors (*Massachusetts History and Social Science Curriculum Framework*, 2003, p. 79). Following the standards, test developers could ask virtually anything about the biographies of these two women (e.g. Sample Test Item 6: "Who wrote *The Feminine Mystique*?"), and they still rightfully claim that it was a standards-based question. Meanwhile teachers, also following the state standards, could discuss Betty Friedan and address her motivations, beliefs and actions, including her role as a founding member of the National Organization for Women, but never require students to read or to know the half dozen or so books that she authored.

The Massachusetts curriculum frameworks, therefore, leave the possibility for a scenario wherein teachers can teach from the standards throughout the academic year, and test makers can devise items derived from those same standards, yet the content the teachers teach and the test items test developers create may not overlap very much. The result may be that the students encounter items on the standards-based test that that their standards-based teachers didn't teach. The implications of this are discussed in detail in the final chapter.

How Teachers Viewed State Standards and the MCAS-US Test

The clear misalignment between HSS teachers and the MCAS-US test does not necessarily mean that HSS teachers believe that the MCAS-US test is not a good assessment. To make that determination, six items were included on the survey of HSS teachers in Massachusetts that addressed teachers' beliefs about the MCAS-US test. As far as can be determined from a detailed literature review, my study represents the first survey of teachers' beliefs about the MCAS-US test. Perhaps related to the novelty of the study, I encountered some obstacles in my efforts to develop a meaningful scale to assess teachers' favorability rating of Massachusetts standards and the MCAS-US test. The six items were derived from a larger national study about teachers beliefs about the impacts of state standards and high stakes testing programs (Abrams, et al., 2003). Although some of those items were more directed to the idea of standards, others were directed toward classroom practice and some are directed to the MCAS-US assessment, they were originally intended to get a general sense of teachers' beliefs about the impact of standards and assessments in their classroom. However, as the data show and as my

argument demonstrates, there are clearly several mitigating factors that prevented teachers from reacting uniformly or predictably to the six items.

Hypothetically, the six items addressing teachers' beliefs about the MCAS-US test should have been strongly and positively correlated with one another if the sixth item ("The state mandated testing program leads some teachers in my school to teach in ways that contradict their own ideas of good education practice") were reverse coded. However, a scale reliability analysis conducted on the six items returned unacceptably low Cronbach's α scores ($\alpha=.606$). The scale score dropped lower when items were collapsed to two possible answers: a) Agree/Strongly agree and b) Disagree/Strongly Disagree ($\alpha=.476$).

To determine if any of the six items held together as a MCAS favorability scale, a principle component factor analysis with oblimin rotation was done. The KMO Measure of sampling adequacy was low at .693 but Bartlett's Test of Sphericity was significant with a value of 171.36. An examination of the pattern matrix shows two components loading with Eigenvalues higher than 1 which explained 54.44% of the total variance between the items. The first component loaded four items. This seemed to be a scale of teachers' fidelity to the state frameworks and the MCAS test within their classroom. Yet, when a scale reliability analysis was done on these four items, the test yielded an unacceptably low Cronbach's α of .613. Meanwhile, the second component loaded three items. These items appeared to be a scale representing teachers' beliefs about the MCAS-US test's ability to assess and improve daily instruction. However, when a scale

reliability analysis was done, Cronbach's α was once again unacceptably low to treat these items as a scale ($\alpha = .480$).

The low Cronbach's α scores for each of these scales reflected some underlying limitations of the six items that were not apparent when the survey was constructed. Individual participants' inconsistent responses to these six items may be a reflection of the fact that the items were vague. For example, some of the items referred to the "state wide mandated test" rather than specifically referring to the MCAS-US test. For that reason, some may have assumed that they were being asked to give their responses to the MCAS program in general rather than to the specific MCAS-US test. The inconsistency in responses may also be a reflection of the fact that HSS teachers vary greatly in their level of familiarity with the multiple aspects of the state frameworks and the MCAS test. Many participants noted on the paper surveys that they either had not read the state frameworks for their grade, or they had not seen the MCAS-US test. Many participants skipped those items; others wrote in that they responded to those items based on earlier ideas about the standards and assessments including what they had heard from their colleagues, rather than direct knowledge. Still others wrote that while they had seen the sample items of the MCAS-US test, they had not seen an actual full-length pilot test. Some said that they had not seen students' scores on the MCAS-US, or had seen them only in the aggregate (i.e. for the whole grade, not for the individual students that they taught). For all of those reasons, participants were not able to make consistent and informed judgments about the MCAS-US test.

While the six items addressing teachers' beliefs about the MCAS-US test did not hold together as a scale, much can be gleaned about teachers beliefs based on the raw data. In this section, I first report teachers' responses to each of the six questions. I next discuss some of the implications of the data. Table 25 presents the proportions of teachers responding to each of the six questions on teachers' beliefs about the MCAS-US test. Responses have been collapsed so that agree and strongly agree are treated as one category and disagree and strongly disagree are treated as one category.

Table 25: Responses to Items Regarding the MCAS-US Test

Item	Agree/ Strongly Agree	Disagree/ Strongly Disagree
The MCAS 10th/11th grade mandated test in US	81.6%	18.4%
History is compatible with my daily instruction		
My district’s curriculum is aligned with the state mandated curriculum.	94.4%	5.6%
Scores on the state-wide mandated test accurately reflect the quality of instruction students have received.	23.6%	76.4%
If I teach to the state standards or frameworks, students will do well on the state-mandated test	64.6%	35.4%
My tests are in the same format as the state mandated test.	44.8%	55.2%
The state mandated testing program leads some teachers in my school to teach in ways that contradict their own ideas of good education practice.	83.0%	17.0%

At first glance, it is apparent why scale reliability scores for these six items were low. Some items that asked similar questions were given widely divergent answers by participants. For example, the first two items ask about alignment between the state

mandated curriculum and the school's curriculum, and the state test's compatibility with daily instruction. Responses indicate that teachers do believe their school's curriculum is aligned with the state curriculum, and their daily instruction is compatible with the MCAS test in United States history. Yet, another item which gets at a similar construct, item four, shows that while teachers think the MCAS is compatible with their daily instruction they have mixed sentiments about the success rates of their students on the MCAS-US test if they teach to the standards or frameworks. Again, while these three items get at the same construct of alignment (i.e. alignment between state frameworks, the MCAS-US-test and teachers' classroom instruction), results show strong variations in teachers' responses to the items.

How could teachers respond in such varying ways to items that get at a similar construct? Two factors may explain this variation in participants' answers to related questions. First, as stated earlier, some teachers were not made aware of performance trends of their students on the pilot administrations of the MCAS-US test. Some teachers may have based their answers about students' success on the test on preconceived beliefs about standardized tests or on what they had heard about students' performance on other MCAS test subjects. They may however, have based their answers on actual students' results. Indeed, many department chairs informally reported discussing trends in students' performance on pilot administrations of the MCAS-US test with faculty in their department.

A second explanation for this inconsistency in answers may be that the MCAS-US test never became a fully operational test. For that reason, teachers may see the test as

“compatible” with their daily instruction in light of the fact that there were no stakes attached to the outcomes of the test. To others, “compatible” may mean that the MCAS-US test has no impact on their daily instruction and therefore they are able to continue on teaching as they did before the MCAS-US test was introduced. Given that participants were split about whether students’ would be successful on the test even if the teacher taught directly from the state frameworks may indicate teachers’ doubts about the quality of the test but a lack of concern about the impact the test would have on their daily instruction in light of the fact that it is not a high-stakes test.

The previous argument is strengthened by the finding that 76.4% of teachers either disagreed or strongly disagreed that students’ scores on the MCAS-US test accurately reflected the quality of instruction the student received. While one would expect that there would be similarities between answers on item three and item four, which asks a similar question about students’ success on the test, enough nuanced difference exists between the two to understand why teachers diverged slightly on these two questions. The biggest difference between items three and four is the insertion of the word “accurately” in item three. Participants could easily believe that if a teachers’ instruction is linked to the frameworks, then students will “do well” on the test, yet the test may not be an “accurate” reflection of teaching quality. Again, this could be related to the fact that the test was not fully operational. Teachers might have felt that they adequately prepared their pupils for the test but that, knowing the test wouldn’t “count” for anything students did not take the test seriously. Many survey respondents wrote notes indicating that because the test was not “high stakes” their students just “blew it

off” or “didn’t even try.” One department chair indicated that when analyzing the tests results for students in the school, she came across one test answer sheet on which the student had apparently used their allotted 45 minute test period to painstakingly shade bubbles to form the picture of an extended middle finger. With results like these, one can hardly question why teachers believe that the test results do not adequately reflect quality of instruction.

Yet, the argument can also be made that teachers don’t believe that students’ scores on the MCAS-US test accurately reflect the quality of instruction students receive because, in their eyes, the MCAS-US test is not a quality assessment. Teachers’ responses to items five and six support that argument. While some have argued that tests can act as a “lever of change” in the classroom in the sense that teachers will reform and improve their curriculum so students will do well on the test (Grant, 2001), less than half of the teachers in this sample reported that their tests were in the same format as the MCAS-US test. In other words, teachers did not feel compelled to alter their pedagogical and assessment technique to align better with the test. Furthermore, an overwhelming majority of teachers (83%) believed that the MCAS-US test leads some teachers in their school to teach in ways that contradict their own ideas of good education practice. Responses from items five and six together suggest that the majority of teachers in this sample believe that the MCAS-US test promotes teaching that they would consider pedagogically unsound.

While the majority of teachers in the survey may believe that the MCAS-US test is unsound, not all teachers in this sample believe this. At the outset of this study, I

theorized that the relationship between teachers and the MCAS-US test might depend on teachers' orientation to the subject. Specifically, I hypothesized that teachers from particular teaching orientations were more likely to support or endorse a standardized history test such as the MCAS-US. To study the relationship between teachers' orientations and their beliefs about the MCAS-US test, I had hoped to create an MCAS-US test favorability scale. I discussed why this was impossible. Regardless, it was possible to look at teachers' responses to each of the six questions about the MCAS-US test based on their calculated teaching orientation. To that end, I conducted a bivariate correlation test to analyze the relationship. My null hypothesis was that there is no correlation between scale scores for any of the five curricular orientations and favorability ratings of the MCAS-US test. My alternative hypothesis was that there is a correlation between scale scores for one or more of the five orientations and favorability ratings of the MCAS-US test. Specifically, I hypothesized that there is a significant and positive correlation between scale scores on the CT and SS orientation and the items that show favorability toward the MCAS-US test. Conversely, I hypothesized that there is a significant and negative correlation between scale scores on the RI, RISE and RIPE orientations and favorability toward that MCAS-US test. That is, I hypothesized that the higher the CT and SS scores were, the more likely they would be to endorse the MCAS-US test. The higher the RI, RISE, and RIPE scores were, the more likely they would be to react negatively to items about the MCAS-US test. Table 26 presents significant results of the bivariate correlation analysis.

Table 26: *Pearson's Correlation between Teachers' Orientation Scale Scores and Beliefs About the MCAS-US Test*

Item	CT Score	RISE Score	RI, Family Score
Item 1: The MCAS 10th/11th grade mandated test in US History is compatible with my daily instruction	.337**	--	--
Item 2: My district's curriculum is aligned with the state mandated curriculum.	--	-.150*	--
Item 3: Scores on the state-wide mandated test accurately reflect the quality of instruction students have received.	.263**	--	--
Item 4: If I teach to the state standards or frameworks, students will do well on the state-mandated test	.202**	--	--
Item 5: My tests are in the same format as the state mandated test.	.274**	--	--
Item 6: The state mandated testing program leads some teachers in my school to teach in ways that contradict their own ideas of good education practice.	--	--	.139*

Note. No significant relationship is indicated by --.

*p < .05 level, two-tailed. **p < .01 level, two-tailed.

Results indicate that there are some significant relationships between teachers' orientations and the MCAS-US test in the direction predicted. Specifically, cultural transmission scale scores were significantly and positively correlated to teachers' responses on several items about their beliefs about the MCAS-US test. That is to say, the higher a teacher scored on the cultural transmission scale, the more likely that teacher was to respond favorably to items one, three, four, and five. However, while the relationship is significant, the magnitude of relationship is very low.

Additionally, RIPE scale scores were significantly and negatively correlated to item number six. Put another way, individuals who scored highly on the RIPE scale were significantly more likely to respond that they either agreed or strongly agreed that "the state mandated testing program leads some teachers in my school to teach in ways that contradict their own ideas of best practices." The results of this bivariate correlation test clearly indicate that the only group that appears to significantly favor (though not strongly) the MCAS test are individuals who scored highly on the cultural transmission scale. This finding makes sense. Having found that the vast majority of sample items on the MCAS test were rated by the four panelists as containing content most favored by cultural transmissionists, it would follow that people who scored high on this scale would agree that the test mirrors their own tests and that results of the MCAS-US test accurately predict high quality instruction in HSS. Meanwhile, it should be noted that individuals whose highest scale score indicated they taught from the cultural transmission orientation were a very small minority of teachers in this sample ($n = 11$ or 4.3%). In other words, the only group that statistically significantly trusted the MCAS-US test to accurately

reflect the work they did in the classroom was the smallest group of Massachusetts HSS teachers.

Based on the data and findings generated by the survey of Massachusetts HSS teachers, a few conclusions can be drawn as to how teachers viewed state standards and the MCAS-US test. First, it appears that teachers have complex, varying and sometimes contradictory opinions about the MCAS-US test. Teachers' beliefs about the MCAS-US test were very difficult to establish and this may be a function of the MCAS-US test's shifting status from "predominately world history" in the late 1990s to "predominately US History" now and from "pilot-test-but-soon-to-be-high-stakes" test in fall 2008 to "suspended-for-the-time-being" test in spring 2009. Teachers most likely have not solidified their stance toward the MCAS-US test. Many of those who do have opinions most likely formed their opinions based on preconceived ideas about the validity of standardized tests in general. Still, many teachers formed their ideas based on observations about the actual format of the MCAS-US test, its connection to state frameworks, its relevance in the classroom, and its ability to generate accurate scores of student achievement.

Overarching trends from the survey analysis seem to suggest that teachers have some misgivings both about what is on the test and how the test will impact classroom instruction. And, while no systematic analysis of open-ended responses to the MCAS scale was conducted it is clear that participants who provided comments in the open-ended response do not favor the MCAS-US test as it is currently formatted. Many question the prominence of multiple choice questions, assessing students based on their

ability to memorize discrete facts that aren't representative of the full history and social studies curriculum. Others write that they wish to see more open-ended response questions such as the document based questions on the New York Regents exam. So, while Massachusetts HSS teachers appear to have significant misgivings about the MCAS-US test, they have many suggestions for amending and improving it. Based on those statements, it seems that while Massachusetts HSS are *not against testing*, they are against *this test*.

Taken together, findings from my content analysis and my survey analysis appear to reinforce one another in a significant way. All teachers, regardless of curricular orientation, indicated on the survey that their district's curriculum was aligned with the state standards. It is likely that the near-unanimous consensus on this item made it difficult to see differentiation on the survey between curricular orientations and to develop an MCAS favorability scale. However, only teachers who aligned with the CT orientation were statistically significantly more likely to believe that if they faithfully taught to the standards, their students would do well on the test and that their students' scores on the test accurately reflected the instruction students were given. This finding is significant especially in light of the fact that teachers who align most with the CT orientation are such a small minority (4.3%) of the Massachusetts HSS teaching force. Moreover, this finding lends further credibility to the suggestion, discussed in the previous section, that test makers' and teachers' varying *interpretations and enactment* of the state standards in their daily work is a significant mitigating factor contributing to the variation in teachers' beliefs about the utility of the MCAS-US test as an assessment

measure. While test makers and teachers who work within the CT orientation interpret and act on the state frameworks in similar ways, teachers who work within the RI orientation (80% of the sampled Massachusetts HSS teaching force) interpret and act on the state frameworks in very different ways. It follows then that test makers and teachers who work within the CT orientation believe the scores on the test accurately reflect the instruction that students receive. Meanwhile, teachers who work within the RI orientation may not believe scores the MCAS-US test accurately reflect their teaching and further, certain teachers who work within this RI orientation (teachers more aligned with the RIPE orientation) believe that the test leads some in their school to teach in ways that are pedagogically unsound.

The findings of this dissertation then lead to many questions about how the differing interpretations of the state frameworks play out, both in terms of teachers' creation of curriculum for their classroom, and in terms of how test developers construct individual test items. Additionally, the findings of this dissertation lead to questions about how teachers and test developers might work in greater unison toward the goal of improving civic knowledge and democratic participation in America's youth. In the next chapter, I review the data, findings, analysis and arguments of this dissertation and consider the implications of this study.

CHAPTER SIX: IMPLICATIONS OF THIS STUDY

This dissertation began with a discussion of the problem of waning civic activity, which seemed to be unaffected by policy interventions for public schools aimed at improving civic knowledge and participation. This is borne out by consistently low test scores on assessments of historical and civic knowledge [e.g. National Assessment of Educational Progress (NAEP)] and waning civic participation. Some blame teachers for failing to teach students meaningful content or important history, and others counter that students' poor performance on tests signal a problem with the test's construction rather than with teachers. A key purpose of this study was to systematically study both the orientations of HSS teachers in the state of Massachusetts and the skills and constructs measured on the MCAS-US history. To study this problem, this dissertation set out to answer three questions:

1. What are the orientations of history and social studies teachers in the state of Massachusetts?
2. What skills and constructs are measured by the proposed high-stakes, state-mandated high school test in United States history?
3. What is the relationship between the orientations of history and social studies teachers in the state of Massachusetts to the constructs measured by the proposed Massachusetts state-mandated history and social studies assessment?

This study considered the complex relationship between HSS teachers' orientations to their subject matter, and the skills and constructs measured on the MCAS-US history test via two research designs. First, a survey of a sample of Massachusetts

HSS teachers was conducted to analyze the orientations from which these teachers approach the subject. Second, a content analysis of the MCAS-US test was conducted to identify the HSS skills and constructs assessed on the test. Both the survey of teachers and the content analysis employed the same three traditions framework for understanding the multiple and varied ways that HSS subject matter is taught and assessed. Additionally, both analyses were informed by the theory of democratic pragmatism.

While the two designs differed, the use of consistent conceptual and theoretical frameworks across both designs allowed for three key comparisons to be made. First, in chapter four, the orientations of Massachusetts HSS teachers were compared to the educational and pedagogical ideals espoused by theorists who work from a democratic pragmatist framework. Second, in chapter five, the skills and constructs measured by the MCAS-US history test were compared to curriculum and assessment ideals consistent with democratic pragmatism. The use in both designs of the three traditions framework allowed for a third comparison in chapter five between the orientations of HSS teachers in Massachusetts and the skills and constructs on the MCAS-US history test. This chapter uses the three comparisons as a means of synthesizing the findings and considering implications. This chapter concludes with suggestions for changes to the MCAS-US test to bring the test closer in line with both democratic pragmatism and the orientations of Massachusetts teachers.

Out of Alignment: The MCAS-US Test and Massachusetts HSS Teachers

This dissertation was designed to inform current controversies about history and social studies education. Here there are debates about whether teachers' orientations and

perspectives, on one hand, or the structure and content of standardized tests, on the other, are responsible for the continued poor performances of students on assessments of civic and historical knowledge. Analysis of the data gathered for this study demonstrates that the curricular orientations of history and social studies teachers in Massachusetts are, indeed, quite different from the skills and constructs that are measured by the proposed state-mandated high school test in United States history. That is to say, HSS teachers and the HSS standardized test appear to have very little in common.

One possible interpretation of the results of this study is that the documented misalignment between HSS teachers' orientations and the content of standardized tests, which are characterized by state authorities as "standards-based," is the fault of teachers who are not faithfully and thoroughly teaching the historic, civic and economic content outlined by the state frameworks. This kind of argument has been made repeatedly in education circles and most prominently by a group whom Banks (1993) refers to generally as "Western traditionalists."

Many Western traditionalists claim that the problem of waning civic knowledge among high school students (and U.S. citizens) reflects reluctance on the part of teachers to directly teach the canonical concepts outlined in the state frameworks. For example, a report commissioned by the Organization of American Historians (2004) emphasized the viewpoint of historian, Sean Wilentz (2003) who claimed that teachers "pose as courageous progressives dedicated to liberating schoolchildren from the tyranny of rote instruction...But if they have their way, the widely lamented historical illiteracy of today's students will only worsen in the generations to come." Along similar lines

Bennett et al. (1990) argued that teachers devise questionable activities such as “pop psychology... self-esteem exercises and propaganda for particular causes” and that these activities “pass for social studies in many schools” (1990). For that reason, Bennett warns, students’ learning in history and social studies above all other major curricular subjects “requires extra vigilance on [the parents’] part” (p. 187).

Further, Western traditionalists often claim that the blame for unacceptably low levels of civic knowledge and participation by youth lies with social studies teacher educators who have created a “thought world...infused with the notion that traditional history and social science content...should be eradicated in order to better prepare young people to reform society” (Leming, et al., 2003, p. ii). This kind of attack on teachers and teacher preparation, which Cochran-Smith, Barnatt, Lahann, Shakman, & Terrell (2008) referred to as “the knowledge critique” of teacher education programs with a social justice theme, is widespread and long-standing. For example, Bennett (1990) wrote almost two decades ago that while a handful of strong teacher preparation programs exist, most are “an embarrassment” (p. 622). Hirsch (1996) contended that an “anti-intellectual, progressive attitude...prevail[s] in the educational community” (p. 48). And, in a more recent *Newsweek* column, George Will (2008) claimed that teacher education programs supplant “rigorous pedagogy” and “teacher-centered classrooms where knowledge is everything” with “vacuity” and a “progressive political catechism” (Cochran-Smith, et al., 2008, pp. 637 - 638). The argument here is that when American students are taught by teachers who were trained to abhor traditional knowledge, students can hardly be

expected to learn foundational historical content, let alone succeed on standardized tests of history.

From this perspective, the solution to the problem of diminishing civic knowledge and participation in America's students is clear. Professional development workshops and teacher education programs (preferably not developed or executed by the teacher preparation community that led teachers astray to begin with) should enable teachers to obtain knowledge more in line with the state frameworks and, in turn, to transmit that knowledge to their students. Those who follow this line of reasoning make several assumptions both about teachers' instruction and about how students learn: a) currently HSS teachers are not teaching canonical, standardized content; b) it is possible for HSS teachers to teach the material outlined in the state frameworks in the instructional time they have with their students; c) if teachers had stronger content knowledge related to history, they could adequately relay that knowledge and students could absorb it well enough to do well on tests of civic knowledge; and d) learning more historical content will lead students to participate in civic life. Each one of these four assumptions is worth critiquing. Specifically, the assumptions about how humans learn that ground this perspective run contrary to current research about how students of history and social studies actually acquire concepts and put that knowledge to use. In fact, much of the research on how HSS students learn provides evidence to suggest that hearing something in class once or even a few times does not result in the construction of knowledge; memorization techniques do not lead to long term acquisition of topics; and exposure to facts does not help students create the broader contextual connections needed for a deep

understanding of history, or lead students to participate in greater numbers in civic life. (Bransford, Brown, & Cocking, 2000; Downey & Levstik, 1991; Levstik & Tyson, 2008)

Contrary to interpreting the results of this study the way Western Traditionalists might, critics of standardized tests might conclude that the study shows that standardized HSS tests, such as the Massachusetts state test, are to blame for the poor performance of students because the tests do not assess the knowledge or skills that teachers focus on in the classroom. This perspective is consistent with the argument that state-mandated tests are based upon unrealistic expectations about how much of the state frameworks teachers can cover in a meaningful way and how much students can absorb.

The arguments of standardized test critics are discussed throughout this study and, similar to the arguments of Western Traditionalists, are widespread. The debate about high stakes tests is rife with accusations about how these tests obscure the true goals of education. The accusations are so often repeated that many assume they are true: “Teachers teach to the test. The test narrows the curriculum. Test scores don’t reflect students’ true abilities. One size can’t fit all.” For example, while the purpose of schooling is hotly debated, social studies teachers often contend that the task of their discipline is to ready pupils for democratic citizenship. Yet test critics claim that standardized tests do not accurately measure pupils’ readiness for this responsibility or measure students’ aptitudes to perform the tasks, skills and logical thinking required of citizenship (Horn, 2006; Kurfman, 1991; Pahl, 2003a; Rothstein, 2004). Rather than validly measuring pupils’ readiness for citizenship, test critics claim the tests measure students’ ability to memorize random, unimportant information.

Still another aspect of standardized assessments in HSS that troubles test critics is the method by which test developers select and develop specific test items from the enormous scope of testable content in the subject delineated by the state standards. Many critics claim that the content test developers choose from is “practically infinite” (Thornton, 2001), and that HSS standards “overemphas[ize] mundane trivia” (Saunders, 1996) and compel teachers to cover everything “from Plato to NATO” (Foster & Morris, 1991). This means that test developers create tests that compel HSS students to commit countless factoids to memory. Even so, there is no guarantee that the facts students memorize will be the same facts that the test assesses. In other words, detractors of standardized tests argue that test items assess very minute and specific data within a practically infinite subject area. Without using specific criteria to guide the selection of content on a test item, there is no way for a teacher to know what content will be tested. For critics of standardized tests, an appropriate correction for the misalignment between teachers and the MCAS-US test would be to end or drastically alter standardized testing.

Outlining the perspectives of both Western traditionalists and test critics serves to elucidate the debate around HSS teaching, standardized testing, and proposed solutions to the enduring problem of low civic knowledge and waning activity. Interestingly, by investigating these arguments, there appears to be one clear common denominator. From the perspective of Western Traditionalists, teachers are to blame for not faithfully following the standards. From the perspective of test critics, standardized tests are to blame for testing some selected portion of the vast array of details included in standards

rather than content that is fundamental to the subject. The common denominator between these two perspectives is state standards.

In chapter five, I pointed out that both teachers and test developers claimed to rely on state standards to inform their work. However, as I argue below, the Massachusetts curriculum frameworks are so broad that it is possible for teachers to teach in ways that are in keeping with the standards throughout the academic year, and it is simultaneously possible for test makers to select and devise items derived from the same standards. And, yet it is also possible that the content teachers teach and the items test developers devise do not overlap much at all. The breadth of the standards along with teachers' and test developers' differing interpretations of those standards make it possible, as my analysis shows, that students are being tested on different content from that which they are taught. When students, in turn, do poorly on the standardized test, it leads many to debate whether tests or teachers are to blame. It might be that the standards are so broadly defined that they do not effectively regulate teaching practices or test development. To return to the metaphor of the three-legged race, while HSS teachers' and test developers' third leg is conjoined by standards, the standards are so broad that teachers and test makers are not working together enough to step in unison. A reasonable compromise to this problem may be for test developers and teachers to streamline the amount of content from the state frameworks that both teachers' curriculum and the test draw on. To do so requires a clear and consistent theory that can be used to guide the selection and prioritization of content from the state frameworks.

In Search of a Guiding Theory for the MCAS-US Test

Test developers often rely on item response theory and other psychometric standards to determine whether items are valid and reliable. They often assess the reliability of an item by verifying that students from different ability levels perform predictably on the given item. In other words, students who perform well or poorly on an item on one instrument should perform well or poorly on a similar item in another instrument (Keller, 2009). Content that is as broad and varied as U.S. history offers many different avenues for test developers to convert content into test items. On any one topic, an item may be phrased in numerous ways to achieve the desired differentiation in the population. Yet, while the item may be statistically good [e.g. students' responses exhibit "spread" (Wineburg, 2004)], it might not be vital to proficiency in the subject.

Determining whether an item tests information that is vital to the subject is an issue of test construction that test developers readily acknowledge as vexing (Dwyer, 2008b; Keller, 2009). This is the case, as they point out, because, even if a student were familiar with an historical theme, a particular test item might not assess the knowledge that is basic to that theme. A possible solution to this vexing problem could be using a unifying theory, such as democratic pragmatism, as a basis for deciding which content should appear on a test. This could help to clarify the test maker's task and lead to a more theoretically sound assessment.

I return to an example introduced in the first chapter of this dissertation to illustrate the problem of test items that assess trivial rather than vital subject knowledge. In *Still at Risk: What Students Don't Know, Even Now*, Hess (2008) made the claim that

teachers were in part to blame for students' poor performance on history assessments. His report gained the public's attention with alarmist statements such as "nearly a quarter of U.S. students could not identify Adolph Hitler" on the *Common Core* assessment (F. Hess, 2008, p. 1). What was not made clear, however, was that even though this statement was accurate in a certain sense, it was misleading. Students might have known about and been able to identify Hitler in myriad ways not asked about on the test. For example, students might know a lot about Hitler's role in the Holocaust. Yet, that was not the basic knowledge that this assessment tested. In this case, the failure of students to identify Hitler was a failure to identify him as "the Chancellor of Germany during World War II." Hypothetically, if a student did not know the details of Hitler's biography, he or she might not have known, for example, whether German leaders are called "Chancellors," "Presidents" or "Prime Ministers." Based on that line of thinking, he or she might have decided that identifying him as "the Chancellor" of Germany was not the correct choice. The item calls on students to recall Hitler's title rather than to think deeply about his evolving role in Germany, the social, economic, and political factors that led to his ascendancy, or his destructive abuses of power. This example calls into question whether Adolph Hitler's title of "Chancellor" is a significant aspect of his biography. In other words, does this question assess, as the *Common Core* report *Still at Risk* claims it does, "basic" history? Rather than assessing something meaningful about German history or Hitler's biography, this question appears to assess whether or not students know that leaders of Germany are called "Chancellor's."

Another issue with HSS standards-based test items is that the standards from which items are derived are often so broad, it is difficult for anyone to predict the specific content within the broad standard that test makers will choose to highlight in a particular test item. For example, the following item (Sample Test Item 11) appeared on the 2006 pilot administration of the Massachusetts high school U.S. History test.

Sample Test Item 11: *Item 5 from the MCAS Guide to History and Social Science Assessments (2007)*

The Battle of Vicksburg was significant because it

- a. marked the end of Confederate invasions into Union territory.*
- b. destroyed the South's greatest city, devastating Confederate morale.*
- c. gave the Union control of the Mississippi and split the Confederacy in two.*
- d. created a moment appropriate for President Lincoln to proclaim Emancipation.*

This question is drawn directly from the Massachusetts framework for history and the social science (2003). The specific standard from which this item was derived states:

USI.39 Analyze the roles and policies of various Civil War leaders and describe the important Civil War battles and events.

Battles

- A. The Massachusetts 54th Regiment and the Battle at Fort Wagner
- B. Antietam
- C. Vicksburg
- D. Gettysburg

The item is “valid” in the sense that it tests content that Massachusetts teachers should cover according to the standards. Students are expected to be able to “describe important Civil War battles” including the Battle of Vicksburg. When this battle is discussed by HSS teachers in class, however, it is likely teachers emphasize it as a Union win and do not necessarily explain the broader implications this victory had for the Civil War. Meanwhile, all of the answer choices for this question assume that the Battle of Vicksburg was a Union victory. For that reason, even students who are familiar with Civil War battles – in that they know who was defeated and who was victorious – might not possess the detailed level of knowledge required to answer this question. It may be unfair to surmise that the student doesn’t know “basic” history when the item doesn’t appear to test basic knowledge.

Another example which brings into sharp relief the infinite amount of testable content in the Massachusetts HSS standards is the following item (Sample Test Item 12).

Sample Test Item 12: *Item 6 from the MCAS Guide to History and Social Science Assessments (2007)*

How did World War II impact the lives of African Americans from 1941 – 1945?

- a. African Americans were widely elected to public office.*
- b. African Americans served with whites in desegregated military units.*
- c. Civil rights reforms greatly improved the social status of African Americans.*
- d. Defense industry jobs led many African Americans to migrate from the South.***

Again, this question is drawn directly from the Massachusetts framework for history and the social sciences (2003). The specific standard from which this item was derived states:

HS.USII.17 Explain important domestic events that took place during the war.

- A. How war-inspired economic growth ended the Great Depression
- B. Philip Randolph and the efforts to eliminate employment discrimination
- C. The entry of large numbers of women into the workforce
- D. The internment of West Coast Japanese-Americans in the U.S. and Canada.

The Second Great Migration is included under this standard because it was an “important domestic event that took place during World War Two” (*Massachusetts History and Social Science Curriculum Framework*, 2003). However, numerous domestic events took place during the war – rationing led to changes in diet and fashion, Victory Gardens were grown, and persistent racial tensions erupted in the Chicago and Detroit race riots and the “Zoot Suit” riots in Los Angeles, to name just a few. A teacher who stringently followed the state standards in United States history would likely focus on the four events that are explicitly delineated by the standards. Considering that the four events would take ample instructional time to cover on their own, it is not so clear that a conscientious teacher would necessarily devote additional time to the Second Great Migration or countless other domestic events that occurred during the war, given that these are not listed in the standards. The Second Great Migration is an important event in United States history which ought to be included in the curriculum. However, the point here is that this test item is drawn from a standard that does not explicitly require teaching this historic event yet an item about the Second Great Migration appeared on the sample test. The standards

are written broadly enough that any test item to do with this time period, regardless of how significant or obscure, is legitimate for the high stakes test.

More specific guidelines are in order and many critics of standardized tests argue that a rule is needed to guide test makers in their decisions about what should and should not be included on these assessments (Gaudelli, 2002; Rothstein, 2004; Wineburg, 2004). Democratic pragmatism could fill the role of a guiding theory because it addresses educational problems that are central to the task of validly and accurately assessing pupils' ability to perform democratic tasks. Those who work from a democratic pragmatist perspective are not relativists when it comes to knowledge and skills; they do not operate from the assumption that all knowledge is equal or that anything goes. Rather, democratic pragmatism offers very clear guidelines about the content and skills citizens in a democracy need to master (Dewey, 1916; Festenstein, 2004; Gutmann, 1987; Gutmann & Thompson, 1996; Hickman & Alexander, 1998). Subject matter should be included in the curriculum if it eases communication and deliberation among citizens. Gutmann (1987) writes that children "learn the three R's largely by direct instruction... [and that] training of this 'didactic' sort is democratically desirable because it enables citizens to understand, to communicate, and in some cases to resolve their disagreements" (p. 50). This means that common culture can serve as a starting point for curriculum (Hickman & Alexander, 1998). But, in addition, competing narratives of American history (multicultural, critical, social, etc.) should be included for the same reason – knowledge of these subjects also eases communication between citizens with very diverse histories.

Indeed, Gutmann (1987) deals at length with the challenges that multiculturalism and nationalism pose for a democratic education. She establishes specific thresholds for content that should be included in democratic education. First, she argues that curricular content should encourage students to recognize the experiences of oppressed groups. The process of democratic deliberation can muffle the history of these groups simply because they are (currently) a minority and democracies favor decisions made by the majority. Teaching about American society without recognizing oppressed groups is “counter-productive to engaging students in learning about the history and politics of their society” (Gutmann, 1987, p. 306). A second threshold for democratic education is instruction about tolerance. Gutmann writes that multicultural societies have diverse conceptions of the “good life.” Rather than endorsing a single conception of the “good life,” she argues that schools must teach toleration of competing conceptions regardless of whether opposing conceptions are justifiable for everyone (p. 308). Gutmann notes that “Any conception of democracy that is committed to treating people as civic equals should defend [multiple conceptions of the ‘good life’]” (p. 305). For all of these reasons, Gutmann appears to endorse the notion that both common culture and competing narratives should be included in the curriculum.

Most importantly, those who work from the perspective of democratic pragmatism argue that the most practical skill a young democrat can learn is the skill of effective inquiry – that is the process of encountering a problem, developing an effective method to investigate the problem, reasoning through potential solutions and, ultimately, finding resolution to the issue. Both the acquisition of concepts and the development of

reasoning skills are key objectives in education, but developing reasoning skills is paramount. Practicing this skill should be at the heart of education (Festenstein, 2004). What I am suggesting here is that the question of what a capable citizen in a democratic society should know and be able to do can be answered, at least in theory, by democratic pragmatists. Test developers could prioritize the development and inclusion of assessment items that provide test takers the opportunity to demonstrate their reasoning skills.

How do these tenets and thresholds established by democratic pragmatists translate to specific guidelines for standardized test developers? They suggest that both taught and tested concepts in HSS should a) serve the purpose of easing communication between citizens by reaffirming pragmatic and useful aspects of common culture, b) include competing narratives (e.g. the experiences of oppressed groups) into the “grand narrative,” c) engender in students a sense of tolerance and appreciation for differences between and among American citizens, and d) reinforce the skills central to methodical inquiry. This would mean that test makers would look at each HSS test item with the question, “Does this qualify as knowledge that an American student *must know* to participate effectively and actively in a diverse democratic society?” Application of these criteria makes many items in the sample MCAS-US test seem out of place, if not somewhat puzzling. Two examples further elucidate this point. The first (Sample Test Item 13) is drawn from the famous Ravitch & Finn (1987) study, *What our 17-Year-Old’s Don’t Know*:

Sample Test Item 13: *Item from Ravitch & Finn (1987), What our 17-Year-Old's Don't Know*

- Which of the following characterizes United States foreign policy during the early 1900s?*
- a. Fifty-four forty or fight?*
 - b. The Good Neighbor policy*
 - c. The business of America is business*
 - d. **Speak softly, and carry a big stick***

Choice “a” refers to a land dispute between the United States and Great Britain; choice “b” refers to President Franklin Roosevelt’s foreign policy; and choice “c” and “d” are quotes made famous by President Calvin Coolidge and President Teddy Roosevelt respectively. All of the choices characterize an aspect of a president’s foreign policy agenda. Is the student who is able to discern the “one best answer” better equipped for American civic life than others who fail to do so? It is unlikely that answering this question correctly legitimately separates prepared from unprepared citizens. In fact, this test item only meets the first of the four criteria listed above for what content should be tested and taught. A second example (Sample Test Item 14) is taken from the 10th/11th grade U.S. History test.

Sample Test Item 14: *Item 33 from the High School US History 2007 Pilot Test*

Released Items

Which of the following issues was central to the Nullification Crisis of 1832 – 1833?

- a. Due process*
- b. Laissez faire*
- c. **States' rights***
- d. Women's rights*

A teacher could hypothetically deal at length with the complicated federalist relationship between states and the national government without mentioning specifically the Nullification Crisis. Aside from demonstrating that a high school student has memorized an historical event, knowing the correct answer says very little about his or her preparedness to participate in a democratic society.

The consequences of failing to answer any of these questions on a high stakes test are severe enough that each of the questions should represent something that students really *must* know to graduate from high school. Test developers and HSS teachers must have a way to choose the historical facts which are most “basic” to the American story or most “important” for a student to learn to participate in American civic life. Applying the four criteria listed above to the item on the Nullification Crisis indicates that, at best, the item addresses the first and the second criteria. The Nullification Crisis is an accepted part of the nation’s common history but also is an example of a time when states (specifically South Carolina) questioned the authority of the national government.

However, the third and fourth criteria, which some say are “critical,” are not addressed by this question. Establishing an explicit set of criteria for what should be included on the MCAS-US test, and judging each item against those criteria is paramount to the task of creating a valid and defensible assessment.

Interestingly, there are a handful of items on the sample MCAS-US test and other released items that demonstrate that the four criteria of democratic pragmatism can be fulfilled by a standardized test item. For example, the following item (Sample Test Item 15) appeared on the spring 2008 pilot test (*High School U.S. History Pilot Test Released Items*, 2008).

Sample Test Item 15: *Item 7 from the High School U.S. History Pilot Test*

Released Items

Write your answer to open-response question 7 in the space provided in your Student Answer Booklet.

The right of citizens to participate in government through voting is an essential part of American democracy. Four amendments to the U.S. Constitution that involve voting rights are listed in the box below.

- *Fifteenth Amendment*
- *Nineteenth Amendment*
- *Twenty-fourth Amendment*
- *Twenty-sixth Amendment*

- a. Describe who has the right to vote in the United States today.*
- b. Explain how voting rights have changed since the U.S. Constitution was ratified in 1788.*
- c. Choose one amendment from the box and explain how it extended the voting rights of American citizens.*

This sample item is both aligned with state standards and meets each of the four criteria described by theorists who work from a democratic pragmatist framework. In light of the fact that the item requires students to be familiar with how amendments to the Constitution apply to current day voting rights, this question serves the purpose of

reaffirming pragmatic and useful aspects of common American culture. The item also allows students to consider the experiences of oppressed groups in the United States by including the Fifteenth Amendment, which following the Civil War recognized all male citizens' (including newly freed African Americans) right to vote, as well as the Nineteenth Amendment, which extended suffrage to women in 1920. This item opens the door for students to consider and/or appreciate the experiences of those who are denied the right to vote today. Finally, the item reinforces the skills central to inquiry by requiring students to use evidence to address an historical and contemporary social problem (the extension of the right to vote) to formulate and express in written form an evidence-based argument.

This sample question and a few other items like it on the sample MCAS-US test demonstrate that history and social studies test items can be written that meet the broad requirements of the state's frameworks and simultaneously meet the more stringent criteria outlined by democratic pragmatism. Those responsible for the MCAS-US test might consider adopting a more explicit and specific conceptual theory for the selection of test items based on the standards. Interestingly, however Massachusetts teachers already appear to be working from a perspective that is consistent with this theory. That is an argument I take up in the next section.

Massachusetts Teachers are Working from a Theory

Throughout this dissertation, I have made multiple references to the fissured nature of history and social studies as an academic subject. Researchers in the field note that the purpose of K-12 history and social studies is vague (Longstreet, 1997-98; Marker

& Mehlinger, 1992), enormous in scope drawing from a number of professional fields (e.g. history, political science, psychology, sociology), and among the most poorly defined of the five academic subjects in the K-12 curriculum (Stodolsky & Grossman, 1995). For all these reasons, teachers act as curricular and instructional “gate keepers” and filter the curriculum depending on their own beliefs about the true purpose of the field (Cornbleth, 1998; Ooka Pang & Gibson, 2001; Thornton, 1991, 2001; Wilson & Wineburg, 1988).

The findings generated from the survey of Massachusetts HSS teachers showed that these HSS teachers overwhelmingly aligned with one orientation to teaching, reflective inquiry. Here the purpose of HSS teaching is to train students in the inquiry skills needed to become effective citizens. Students are expected to build upon their knowledge of traditional history and social studies concepts to critically assess their world. Teachers in the RI tradition often treat controversial issues or problems in contemporary society as “starting points” for the curriculum (Janzen, 1995). These problems may be issues that students choose to research (Brubaker, et al., 1977; Evans, 1990; Janzen, 1995; Vansledright & Grant, 1994) but are sometimes problems chosen by the teacher. Thornton (1994) describes the reflective inquiry orientation as “transformative” rather than “mimetic” in that rather than arriving at predetermined answers to questions, students are expected to construct their own solutions to problems. Based on the data gathered for this study, teachers in Massachusetts view this reflective inquiry approach to teaching as a means to a much larger end. That is, teachers expect that their students will use their growing understanding of local, national and

international issues to incentivize students to develop personally, or to become agents of social change.

In chapter four, I developed a lengthy argument to support the contention that the reflective inquiry curricular approach, embraced by the vast majority of sampled Massachusetts HSS teachers, is consistent with a theory of democratic pragmatism. Indeed, from the framework of democratic pragmatism, the most practical skill a young democrat can learn is the skill of effective inquiry – the process of encountering a problem, developing an effective method to investigate the problem, reasoning through potential solutions and, ultimately, finding resolution to the issue. Democratic societies require citizens who are adept at inquiry because effective deliberation, including debate, discussion and persuasion that occur over a problem, are central to the democratic process. Consequently, teachers in a democratic society must nurture and strengthen these skills in their students by having them *practice* an inquiry approach to resolving real world problems. Meanwhile, the same skills involved in methodical inquiry are central to the task of teachers working within the reflective inquiry orientation – as central as the name affiliated with this orientation suggests. It can therefore be concluded that the vast majority of teachers in Massachusetts, as described in chapter four, reported that their teaching was consistent with a perspective of democratic deliberation and civic participation.

Those working on Massachusetts education policy, therefore, have an advantage over national education policy workers in that the HSS teachers who work in the state present a remarkably uniform image about what history and social studies teaching

should look like. The fact that the vast majority of Massachusetts teachers respond in a similar way to a survey that gets at their teaching orientations has advantages. State education policy experts could craft education policy to work in accord both with that theory and with orientations of the Massachusetts HSS population. Specific ways that this can happen are spelled out in the next section.

Out of Alignment: Should the Test or Teachers come in Line?

It is hardly surprising that various education advocates – all of whom presumably are committed to developing successful citizens – could promote widely divergent means to achieve the same end. In fact, history and social science education professionals have noted that those who teach the subject, those who educate teachers who will then teach the subject, and those who test pupils’ knowledge in the subject often work without a uniform theory guiding decisions about how to proceed, what to include and what the ultimate aim should be (Popkewitz & St. Maurice, 1991). Put another way, there appears to be an absence of consensus within the field about the desired outcomes of history and social studies. Classroom teachers, state standards and test developers, and state policy officials must consistently check their actions with clear criteria that serve to bind their work. To give students a fair shot at achieving their highest potential in the subject, on the test, and on the democratic stage, education leaders must be linked by a common theoretical mission. A failure to do so cheats students.

Democratic pragmatism offers a theory about what should be taught and why it should be taught. Using democratic pragmatism as a measure by which we can understand and evaluate the “content message” of the MCAS-US test provides a

surprisingly clear picture. The sample items of the MCAS-US test primarily assess traditional knowledge, and the skills most tapped by the multiple choice items are fact recall and basic reasoning rather than complex reasoning, extended reasoning or deliberation. Given that, it seems that the test promotes a type of historical thinking that is out-of-step with what democratic pragmatists and Reflective Inquirer practitioners' value. For all of these reasons, it might be appropriate to rethink the MCAS-US test and perhaps create a new test that reflects better both the professional aims of history and social studies teachers and the theoretical aims elucidated by those who work from a democratic pragmatist frame.

Though the MCAS-US test is not yet operational, state education officials cannot decide to end the test without changing Massachusetts law, and state education officials plan to pilot the test again beginning in spring 2011. The test in its current iteration requires that students have committed to memory vast amounts of historical data – much of which is not tied together conceptually. Additionally, the survey results gathered from this sample of Massachusetts HSS teachers demonstrate that teachers have misgivings about the test in its current form. Finally, many critics have argued that traditional history tests such as the MCAS-US do not assess students' true proficiency with the concepts and skills most required of citizens in a democracy – that is to say, the test does not require students to demonstrate complex and extended reasoning or an ability to systematically inquire into, deliberate over, or defend a solution to authentic social problems. Test developers, then, appear to have a three-pronged challenge on their hands.

First, the sheer breadth of content in HSS may require that policymakers investigate more precise and discriminating ways to make decisions about what items appear on the test, and that these decisions are based on some comprehensive guiding theory about the purpose of history and social studies education. I have argued that democratic pragmatism can fill that role. Reflective inquiry as a teaching tradition is most in line with the ideals promoted by democratic pragmatism. These teachers encourage students to think about the world outside of the world they know. They work to develop their students' skills of critical investigation and reasoned deliberation. They encourage the ethics of tolerance for diverse opinions, appreciation for multiple world views. These skills are central to democratic deliberation and citizenship and are the skills that democratic pragmatism endorses.

Furthermore, the few items that exist on the sample MCAS-US test that were designated as reflective inquiry content have the clearest connection to the guidelines set by democratic pragmatism. Test items that were designated as reflective inquiry content assess students' ability to explore new information, weigh competing arguments and make informed decisions based on existing evidence. These skills that are crucial to democratic deliberation and citizenship are overshadowed when tests are created that promote a shallow emphasis on fact recall.

Underlying the MCAS-US test is an unstated theory that favors student mastery of seemingly limitless content. Many critics, however, including the Western Traditionalists discussed above, make the claim that mastery over this type of knowledge does indeed prepare students for citizenship (Hirsch Jr., 1996; Ravitch & Finn, 1987; Stotsky, 2004).

There is no doubt that democratic pragmatism would endorse the notion of students being taught common history. However, democratic pragmatism sheds serious doubt on the stance that curriculum and assessment should begin and end with the mastery of common history.

Second, state policy makers and test developers might consider HSS teachers' orientations as they develop the standardized test. For example, they could research the many ways in which teachers transform standards into classroom practice. Doing so would help test developers isolate which curriculum standards are highlighted and which standards are "covered" but not in the way that test developers presume. That information could be used to create a stronger assessment of what students are taught and what they learn. The fact that HSS teachers in Massachusetts seem overwhelmingly to be aligned with a particular teaching orientation should make the task of investigating classroom practice easier – sampled HSS teachers painted a very clear and uniform picture of what HSS education and assessment *should* look like in their responses to the survey. Test developers might, therefore, create a test that works with, rather than against, the skills and concepts teachers emphasize.

Test developers could also invite teachers earlier, and in more meaningful ways, into the test construction process. Massachusetts state-level test developers have included teachers in this conversation only to a restricted degree. Massachusetts HSS teachers were asked to volunteer for Assessment Development Committees (ADCs). Yet, the format of the MCAS-US test was largely determined before teachers were included in these committees (Porter, 2009). Many of the suggestions that teachers in this study gave

for improving the MCAS-US test would have been out of the question at that stage of development.

Many teachers in history and social studies readily recognize the bind of high stakes standardized tests. With high stakes tests increasingly the norm as a graduation requirement in academic subjects including math, English/language arts and science, history and social studies are likely to be “squeezed out” if this subject does not also find a place of its own within the agenda (Burroughs, et al., 2005; Grant, 2006; Horn, 2006; Kurfman, 1991; Savage, 2003; Volger, 2003). HSS teachers therefore recognize both the importance and the consequences of high stakes tests. Indeed, this study found that many HSS teachers were not against testing, per se, although they did have significant misgivings about the MCAS-US test in its current iteration. But many of these teachers were willing to consider a revised version of the test, which is discussed below.

Finally, test developers might consider devising an assessment that requires students to demonstrate that they are able to do the real work of democratic citizenship. Test makers might consider whether success on the test conveys something meaningful about a student’s ability to do more than memorize discrete facts. A good test would require students to express the depths of their knowledge and skills in the subject. Test developers can do this while still upholding the same standards of excellence in test writing that they have worked toward in past iterations of the test. This, needless to say, is a daunting task. For example, a revised test might require students to demonstrate their ability to inquire into social problems, to gather evidence and to craft an argument based on that evidence. This would require students to show that they can rationally deliberate

with those whom they disagree, that they can weigh multiple interpretations of an issue and chose a stance based both on the facts as well as their own values.

Test developers in Massachusetts would not have to devise these tests from scratch. Indeed, many teachers in this sample wrote that they would favor a standardized assessment more akin to the state-wide assessments like the New York Regents and the Connecticut Academic Performance Test (CAPT). For these tests, students are presented with multiple sources from different perspectives regarding a specific social problem. Students are then asked to choose a standpoint, gather evidence from each of the sources and write an argument in defense of their perspective. Yeh (2002) argues that these are the types of tests that are “worth teaching to” because they “allow teachers to focus on teaching critical thinking rather than the universe of items that students might otherwise be asked to recall” (p. 12). This form of testing is much more in line with the theories of education promoted by democratic pragmatists, as well as the pedagogical and curricular aspirations of reflective inquirers.

Furthermore, success on this type of test might provide evidence that students possess the skills and knowledge to succeed as a citizen in 21st century America. In the early 19th century, Alexis de Tocqueville (2000) admired Americans’ ability to discuss controversial political issues with those whom they disagreed, and he surmised that their willingness to do so resulted in an open, tolerant and flexible democracy (p. 491). In sharp contrast to Tocqueville’s 19th century observations, in his book *Bowling Alone* (2000), Robert Putnam lamented the waning level of day-to-day civic exchanges between America’s citizens including the decrease in the amount of time individuals spent

discussing politics with friends and acquaintances. Without the daily political exchanges that allowed people to deliberate issues with those whom they disagree, Putnam asserted, it became easy for people to “demonize anyone who disagree[d]” (Putnam, 2000, 342). In light of these observations, rather than creating an assessment that solely tests students’ factual knowledge, it seems prudent to also assess students’ skills of democratic deliberation including their ability to employ reason to dissect competing viewpoints. These social skills are central to an enduring democracy and to success in 21st century life.

Many base their case for high stakes standardized tests in U.S. history on the fact that citizens don’t have the background knowledge or the civic disposition to take action for the public good. Yet, one has to question whether the types of multiple choice questions included on the MCAS-US test represent the knowledge citizens must have to act in the public’s interest. Will teaching students about American’s common and traditional history help in the endeavor to bring about a greater sense of fraternity in students? Perhaps. Will that same type of historical and social education allow students to appreciate diverse perspectives, to negotiate conflicts about controversial issues, or to learn how to take a stand for what they believe? Probably not.

The problem of diminishing social capital and historical knowledge is not an imagined one. If Americans are indeed worried about problems such as declining rates of participation and civic mindedness, humanity’s diminishing sense of empathy and fraternity, or students’ abilities to make reasoned and well-informed decisions, the sights must be set high to remedy these exceptional problems. More importantly, they must

devise a solution that will bring about the desired change. This study raises the issue that test advocates and developers appear to have more on their hands than creating a test that is psychometrically sound and aligned with state-wide content standards. They also need to address the issue of the test's misalignment with the curriculum that is enacted by teachers in the classroom. At the very least, test makers and teachers who prepare students to take the test should not be working at cross purposes – as seems to be the case in Massachusetts.

Most importantly, test developers must devise a meaningful assessment with the goal of measuring what Americans value, rather than asking Americans to value what the test measures. Knowing that Massachusetts residents have historically welcomed – and in some cases provoked – a good fight, it is unlikely that they will keep their disdain for mismatched educational policies under wraps. The long and tumultuous history of the MCAS test provides evidence of the willingness of Massachusetts residents to question policies that they do not support. This poor track record may lead some to surmise that history and social studies education cannot be meaningfully or accurately assessed and others to resolutely defend the test in its current iteration. However, as with most complex problems, the solution requires time, deliberation and compromise. Education policy, including educational assessments that are motivated by a sound and sincere desire to improve society, should be tenaciously pursued. While it may take more time to create an assessment that the public, educational leaders and classroom teachers can endorse, this does not diminish the virtue of the goal.

BIBLIOGRAPHY

- Abowitz, K. K., & Harnish, J. (2006). Contemporary Discourses of Citizenship. *Review of Educational Research, 76*(4), 653 - 690.
- Abrams, L. M., Pedulla, J. J., & Madaus, G. F. (2003). Views from the Classroom: Teachers' Opinions of Statewide Testing Programs. *Theory into Practice, 42*(1), 18 - 29.
- Ananda, S. (2003). *Rethinking issues of alignment under No Child Left Behind*. San Francisco: WestEd.
- Anderson, C., Avery, P. G., Pederson, P. V., Smith, E. S., & Sullivan, J. L. (1997). Divergent perspectives on citizenship education: A Q-method study and survey of social studies teachers. *American Educational Research Journal, 34*(2), 333 - 364.
- Andres, P. N. (1982). Social studies orientations of secondary classroom teachers in the state of Indiana. Unpublished Dissertation. Indiana University.
- Banks, J. (1993). The canon debate, knowledge construction and multicultural education. *Educational Researcher, 22*(5), 4 - 14.
- Banks, J. A. (2004). Multicultural education: Historical development, dimensions and practice. In J. A. Banks (Ed.), *Handbook of research on multicultural education* (2nd ed., pp. 3-29). San Francisco: Jossey-Bass.
- Barr, R. D., Barth, J. L., & Shermis, S. S. (1977). *Defining the social studies*: National Council for the Social Studies.
- Barr, R. D., Barth, J. L., & Shermis, S. S. (1978). *The nature of the social studies*. Palm Springs: ETC Publications.
- Barth, J. L. (1997). Beliefs the discipline the social studies. *International Journal of Social Education*.

- Bazerman, C. (2006). Analyzing the Multidimensionality of Texts in Education. In J. Green, G. Camilli & P. B. Elmore (Eds.), *Handbook of Complementary Methods in Education Research* (pp. 141 - 159). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Bennett, C., & Spalding, E. (1992). Teaching the social studies: Multiple approaches for multiple perspectives. *Theory and Research in Social Education, 20*(3), 263 - 292.
- Bennett, W., Finn Jr, C., & Cribb Jr, J. (1990). *The Educated Child: A Parents Guide From Preschool Through Eighth Grade*. New York: Touchstone.
- Bergman, M. (2005). *U.S. Voter Turnout Up in 2004, Census Bureau Reports*. Retrieved from <http://www.census.gov/Press-Release/www/releases/archives/voting/004986.html>.
- Bhola, D. S., Impara, J. C., & Buckendahl, C. W. (2003). Aligning tests with states' content standards: Methods and issues. *Educational Measurement: Issues and Practice, 22*(3), 21 - 29.
- Bliss, D., & Banks, D. N. (1994). *Characteristics of secondary social studies teachers: An update of Ochoa's 1981 report* (Reports -- Research No. ED409240).
- Booher-Jennings, J. (2005). Below the bubble: "Educational triage" and the Texas accountability system. *American Educational Research Journal, 42*(2), 231 - 268.
- Bransford, J. D., Brown, A. L., & Cocking, R. R. (2000). *How people learn: Brain, mind, experience, and school*. Washington, D.C.: National Academy Press.
- Brown, S. D. (2006). History teacher certification standards in the states. *The History Teacher, 39*(3), 367 - 380.
- Brubaker, D. L., Simon, L. H., & Williams, J. W. (1977). A conceptual framework for social studies curriculum and instruction. *Social Education, 41*(3), 201 - 205.

- Bryk, A. S., & Schneider, B. (1996). Social Trust: A Moral Resource for School Improvement. In G. G. Wehlage & J. A. White (Eds.), *Rebuilding the Village: Social Capital and Education in America* (pp. 1 - 32). London: Falmer Press.
- Buckles, S., Schug, M., & Watts, M. (2001). A national survey of state assessment practices in the social studies. *The Social Studies*, 92(4), 141 - 146.
- Burroughs, S. (2002). Testy Times for Social Studies. *Social Education*, 66(5), 315 - 319.
- Burroughs, S., Groce, E., & Webeck, M. L. (2005). Social studies education in the age of testing and accountability. *Educational Measurement: Issues and Practice*, 24(3), 13 - 20.
- Campbell, D. (2001). Bowling Together *Education Next*, 1(3), 55 - 61.
- Cimbricz, S. (2002). State-Mandated Testing and Teachers' Beliefs and Practice. *Education Policy Analysis Archives*, 10(2).
- Clandinin, D., & Connelly, F. (1995). Teachers' professional knowledge landscapes: Secret, shared and cover stories. In D. Clandinin & F. Connelly (Eds.), *Teachers' Professional Knowledge and Landscapes* (pp. 1 - 15). New York, NY: Teacher College Press.
- Cochran-Smith, M., Barnatt, J., Lahann, R., Shakman, K., & Terrell, D. (2008). Teacher Education for Social Justice: Critiquing the Critiques. In W. Ayers, T. Quinn & D. Stovall (Eds.), *Handbook of Social Justice in Education* (pp. 625 - 639). New York: Routledge.
- Cohen, S. (2008). What form should the test take? Disagreements over Assessments in Massachusetts.
- Cornbleth, C. (1982). Citizenship Education. In E. Mitzel (Ed.), *Encyclopedia of educational research* (pp. 259 - 265). New York: Macmillan.
- Cornbleth, C. (1985). Reconsidering social studies curriculum. *Theory and Research in Social Education*, 13(2), 31 - 45.

- Cornbleth, C. (1998). An America Curriculum? *Teachers College Record*, 99(4), 622 - 646.
- Cornbleth, C. (2006). Curriculum and students: Diverting the public interest. In G. Ladson Billings & W. F. Tate (Eds.), *Education Research in the Public Interest: Social justice, action and policy* (pp. 274). New York: Teachers College Press.
- Council of Chief State School Officers. (2002). *Models for alignment analysis and assistance to states*. Washington, DC.
- Dewey, J. (1916). *Democracy and education: An introduction to the philosophy of education*. New York: The Free Press.
- Dewey, J. (1944). *Democracy and education: An introduction to the philosophy of education*. New York: The Free Press.
- Dillman, D. A. (2000). *Mail and internet surveys: The tailored design method* (Second ed.). New York: John Wiley & Sons, Inc.
- Doppen, F. H. (2006/2007). Pre-service social studies teachers' perceptions of high-standards, high-stakes. *International Journal of Social Education*, 21(2), 18 - 45.
- Downey, M., & Levstik, L. (1991). Teaching and Learning History. In J. Shaver (Ed.), *Handbook of Research on Social Studies Teaching and Learning: A Project of the National Council for the Social Studies*. New York: Macmillan Publishing.
- Dwyer, K. M. (2008a). Phone Conversation with Kevin Dwyer. In Dianna Terrell (Ed.) (pp. Telephone Conversation). Nashua, NH.
- Dwyer, K. M. (2008b). Telephone Conversation. In Dianna Terrell (Ed.). Nashua, NH.
- Dwyer, K. M. (2009). History MCAS delay may be considered (Globe Story). In D. Terrell (Ed.). Nashua, NH.
- Eisenhart, M., Shrum, J. L., Harding, J. R., & Cuthbert, A. M. (1988). Teacher beliefs: Definitions, findings and directions. *Educational Policy*, 2(1), 51 - 70.

- Evans, R. W. (1988). Lessons from history: Teacher and student conceptions of the meaning of history. *Theory and Research in Social Education, 16*(3), 203 - 225.
- Evans, R. W. (1989). Teacher conceptions of history. *Theory and Research in Social Education, 17*, 210 - 240.
- Evans, R. W. (1990). Teacher conceptions of history revisited: Ideology, curriculum, and student belief. *Theory and Research in Social Education, 18*(2), 101 - 138.
- Evans, R. W. (2006). The social studies wars, now and then. *Social Education, 70*(5), 317 - 321.
- Evans, R. W., & Pang, V. O. (1995). National Standards for United States History: The storm of controversy continues. *The Social Studies, 86*(6), 270.
- Feldman, D. (2007). Citizenship education: Current perspectives from teachers in three states. *Educational Research Quarterly, 30*(4), 3 - 15.
- Festenstein, M. (2004). Deliberative democracy and two models of pragmatism. *European Journal of Social Theory, 7*(3), 291 - 306.
- Fickel, L. H. (2000). Democracy is messy: Exploring the personal theories of a high school social studies teacher. *Theory and Research in Social Education, 28*(3), 359 -
- Fickel, L. H. (2006). Paradox of practice: Expanding and contracting curriculum in a high-stakes environment. In S. G. Grant (Ed.), *Measuring History: Cases of state-level testing across the United States* (pp. 338). Greenwich, CT: Information Age Publishing.
- Ford, P. L. (Ed.). (1899). *The Writings of Thomas Jefferson* (Vol. 10). New York: Putman's Sons.
- Foster, S., & Morris, J. W. (1991). Full of sound and fury...signifying nothing? Why the history standards have a troubled future. *International Journal of Social Education, 6*(2), 60 - 68.

- Fraenkel, J. R., & Wallen, N. E. (1991). Quantitative research in social studies education. In J. P. Shaver (Ed.), *Handbook of Research on Social Studies Teaching and Learning*. New York: Macmillan Publishing Company.
- Fraenkel, J. R., & Wallen, N. E. (2006). *How to design and evaluate research in education* (6th ed.). Boston: McGraw Hill.
- Gagnon, P. (2003). *Educating democracy: State standards to ensure a civic core*. Washington, D.C.: The Albert Shanker Institute.
- Gaudelli, W. (2002). U.S. kids don't know U.S. history: The NAEP study, perspectives, and presuppositions. *The Social Studies*, 93(5), 197 - 201.
- Gay, G. (1997). The relationship between multicultural and democratic education. *The Social Studies*, 88(1), 5 - 11.
- Gay, G., & Boyle-Baise, M. (1996). Multicultural social studies: Ideology and practice. *The Social Studies*, 87(2), 81.
- Goodman, J., & Adler, S. (1985). Becoming an elementary social studies teacher: A study of perspectives. *Theory and Research in Social Education*, 13(2), 1 - 20.
- Gradwell, J. M. (2006). Teaching in spite of, rather than because of, the test: A case of ambitious history teaching in New York state. In S. G. Grant (Ed.), *Measuring History: Cases of state-level testing across the United States* (pp. 338). Greenwich, CT: Information Age Publishing.
- Grant, S. G. (1996). Locating authority over content and pedagogy: Cross-current influences on teachers' thinking and practice. *Theory and Research in Social Education*, 24(3), 237 - 272.
- Grant, S. G. (2000). Teachers and tests: Exploring teachers' perceptions of changes in the New York state testing program. *Education Policy Analysis Archives*, 8(14).

- Grant, S. G. (2001). An uncertain lever: Exploring the influence of state-level testing in New York state on teaching social studies. *Teachers College Record*, 103(3), 398 - 426.
- Grant, S. G. (2007). High-Stakes Testing: How are social studies teachers responding? *Social Education*, 71(5), 250 - 254.
- Grant, S. G. (Ed.). (2006). *Measuring history: Cases of state-level testing across the United States*. Greenwich, CT: Information Age Publishing.
- Grant, S. G., & Vansledright, B. (1996). The dubious connection: Citizenship education and the social studies. *The Social Studies*, 87(2), 56 - 59.
- Grelle, B., & Metzger, D. (1996). Beyond Socialization and Multiculturalism: Rethinking the task of citizenship education in a pluralistic society. *Social Education*, 60(3), 147 - 151.
- Gutman, A. (1982). What's the use of going to school? The problem of education in utilitarian and rights theories. In A. Sen & B. Williams (Eds.), *Utilitarianism and beyond*. Cambridge: University Press.
- Gutmann, A. (1987). *Democratic education*. Princeton: Princeton University Press.
- Gutmann, A., & Thompson, D. (1996). *Democracy and disagreement: Why moral conflict cannot be avoided in politics, and what should be done about it*. Cambridge: The Belknap Press of Harvard University.
- Henry, M. S. (1987). The Advanced Placement American History Examination: How has it changed? *Social Studies*, 78(4), 159 - 162.
- Hess, D. E. (2005). How do teachers' political views influence teaching about controversial issues? *Social Education*, 69(1), 47 - 48.
- Hess, D. E. (2008). Controversial Issues and Democratic Discourse. In L. S. Levstik & C. A. Tyson (Eds.), *Handbook of Research in Social Studies Education*. New York: Routledge.

- Hess, F. (2008). *Still at risk: What students don't know, even now*. Washington, DC: Common Core.
- Hickman, L. A., & Alexander, T. M. (Eds.). (1998). *The essential Dewey: Pragmatism, Education, Democracy* (Vol. I). Bloomington: Indiana University Press.
- High School U.S. History Pilot Test Released Items*. (2008). Malden, MA: Massachusetts Department of Elementary and Secondary Education.
- High School U.S. History Released Items*. (2007). Malden, MA: Massachusetts Department of Elementary and Secondary Education.
- Hirsch Jr., E. D. (1996). *The Schools We Need and Why We Don't Have Them*. New York: Doubleday.
- Hoge, J. D. (2002). Character education, citizenship education, and the social studies. *The Social Studies*, 93(3), 103 - 108.
- Horn, C. (2006). The technical realities of measuring history. In S. G. Grant (Ed.), *Measuring History: Cases of state-level testing across the United States* (pp. 338). Greenwich, CT: Information Age Publishing.
- Hunt, G. (Ed.). (1900 - 1910). *The Writings of James Madison*. New York: G. P. Putnam's Sons.
- Jacoby, J. (2009, March 18, 2009). Cradle of Democracy? *The Boston Globe*.
- Janzen, R. (1995). The social studies conceptual dilemma: Six contemporary approaches. *The Social Studies*, 86(3), 134 - 140.
- Johnston, G., & Percy-Smith, J. (2003). In search of social capital. *Policy and Politics*, 31(3), 321-334.
- Johnston, M. (1990). Teachers' backgrounds and beliefs: Influences on learning to teach in the social studies. *Theory and Research in Social Education*, 18(3), 207 - 233.

- Kaltsounis, T. (1997). Multicultural education and citizenship education at a crossroads: Searching for common ground. *The Social Studies*, 88(1), 18 - 22.
- Kauffman, J. M. (2002). *Education deform: Bright people sometimes say stupid things about education*. Lanham, MD: The Scarecrow Press, Inc.
- Keller, J. (2007). *The Bluest State: How Democats Created the Massachusetts Blueprint for Political Disaster*. New York: St. Martin's Press.
- Keller, L. (2009). *Operationalizing Validity: Challenges and issues encountered in large scale assessments*. Paper presented at the New England Educational Research Organization (NEERO).
- Kelly, T., Meuwissen, K., & Vansledright, B. (2007). What of history? Historical knowledge within a system of standards and accountability. *International Journal of Social Education*, 22(1), 115 - 145.
- Kliebard, H. M. (2004). *The struggle for the American curriculum: 1893 - 1958* (Third ed.). New York: Routledge Falmer.
- Knight, J., & Johnson, J. (1996). Political consequences of pragmatism. *Political Theory*, 24(1), 68 - 96.
- Kohut, A. (2007). *Public Knowledge of Current Affairs Little Changed by News and Information Revolutions. What Americans Know: 1989-2007*. Washington: The Pew Research Center.
- Kornblith, G. J., & Lasser, C. (2004). "Will that be on the exam?" The role of testing in teaching and learning American history. *The Journal of American History*, 90(4).
- Kurfman, D. G. (1991). Testing as a context for social education. In J. P. Shaver (Ed.), *Handbook of research on social studies teaching and learning: A project of the National Council for the Social Studies* (pp. 310 - 320). New York: Macmillan Publishing Company.

- Landman, J. (2000). A state-mandated curriculum, a high-stakes test: One Massachusetts high school history department's response to a very new policy context. Unpublished Qualifying Paper. Harvard Graduate School of Education.
- Lee, J., & Weiss, A. (2007). *The nation's report card: U.S. History 2006 (NCES 2007 - 474)* (No. (NCES 2007 - 474)). Washington, D.C.: U.S. Government Printing Office.
- Leming, J. S. (1991). Teacher characteristics and social studies education. In J. P. Shaver (Ed.), *Handbook of research on social studies teaching and learning* (pp. 661). New York: McMillan Publishing Company.
- Leming, J. S. (1992). Ideological perspectives within the social studies profession: An empirical examination of the "Two Cultures" thesis. *Theory and Research in Social Education, 20*(3), 293 - 312.
- Leming, J. S., Ellington, L., & Porter-Magee, K. (2003). *Where did social studies go wrong?* Washington, D.C.: Thomas B. Fordham Institute.
- Leming, J. S., Ellington, L., & Schug, M. (2006). The state of social studies: A national random survey of elementary and middle school social studies teachers. *Social Education, 70*(5), 322 - 327.
- Levstik, L. S. (2000). Articulating the Silences: Teachers' and Adolescents' Conceptions of Historical Significance. In Peter N. Stearns, P. Seixas & S. Wineburg (Eds.), *Knowing, Teaching, and Learning History: National and International Perspectives*. New York: New York University Press.
- Levstik, L. S., & Tyson, C. A. (2008). *Handbook of Research in Social Studies Education* (First ed.). New York: Routledge.
- Lobes, L. S. (1998). Surveying state standards: National History Education Network's 1997 Report on state social studies standards. *The History Teacher, 31*(2), 221 - 234.
- Longstreet, W. S. (1997-98). Reflections on a discipline of the social studies. *International Journal of Social Education, 12*(2), 25 - 32.

- Lortie, D. (1975). *School teacher: A sociological study*. Chicago: University of Chicago Press.
- MacGilvray, E. A. (1999). Experience as experiment: Some consequences of pragmatism for democratic theory. *American Journal of Political Science*, 43(2), 542 - 565.
- Madaus, G. F. (1988). The influence of testing on the curriculum. In L. Tanner (Ed.), *Critical issues in curriculum: Eighty Seventh Yearbook of the National Society for the Study of Education* (pp. 83 - 121). Chicago: University of Chicago Press.
- Manzo, K. K. (2002, May 15, 2002). U.S. history again stumps senior class. *Education Week*, p. 2.
- Marker, G., & Mehlinger, H. (1992). Social Studies. In P. W. Jackson (Ed.), *Handbook of research on curriculum: A project of the American Educational Research Association*. New York: MacMillan Publishing Company.
- Martínez Alemán, A. (2001). The ethics of democracy: Individuality and educational policy. *Educational Policy*, 15(3), 379-403.
- Martorella, P. H. (2001). *Teaching social studies in middle and secondary schools* (3rd ed.). Upper Saddle River, N.J.: Merrill Prentice Hall.
- Massachusetts History and Social Science Curriculum Framework*. (2003). Malden: Massachusetts Department of Elementary and Secondary Education.
- Mathison, S. (1988). The perceived effects of standardized testing on teaching and curricula. 49(1). Dissertation Abstracts International.
- MCAS Guide to History and Social Science Assessments*. (2007). Medford, MA: Massachusetts Department of Education: Malden, MA.
- Metzger, D. J., & Marker, P. M. (1992). Teaching for democracy: An agenda for social studies teacher education in the twenty-first century. *The International Journal of Social Education*.

- Mitzel, H. E. (Ed.). (1982). *Encyclopedia of educational research/sponsored by the American Educational Research Association* (5th ed.). New York: Free Press.
- Morrissett, I. (1977). Curriculum Information Network Sixth Report: Preferred approaches to the teaching of social studies. *Social Education, 41*(3), 206 - 209.
- Morrissett, I., & Haas, J. D. (1982). *Rationales, goals, and objectives in social studies*. Boulder, CO: Social Science Education Consortium.
- Myers, C. B., Adler, S., & Brandhorst, A. (2002). *National standards for social studies teachers*. Silver Spring, MD: National Council for the Social Studies.
- Nash, G. B., Crabtree, C., & Dunn, R. E. (2000). *History on trial: Culture wars and the teaching of the past*. New York: Vintage Books: A Division of Random House.
- National Commission on Excellence in Education. (1983). *A nation at risk: The imperative for educational reform*. Washington, D.C.: National Commission on Excellence in Education.
- Nelson, M. R. (1998). Are teachers stupid? -- Setting and meeting standards in social studies. *The Social Studies, 89*(2), 66 - 70.
- Ochoa, A. S. (1981). A profile of social studies teachers. *Social Education, 45*(6), 401 - 404.
- Olson, L. (2003). Standards and tests: Keeping them aligned. *Research Points: Essential Information for Educational Policy, 1*(1).
- Ooka Pang, V., & Gibson, R. (2001). Concepts of democracy and citizenship: Views of African American teachers. *The Social Studies, 92*(6), 260 - 266.
- Organization of American Historians. (2004). History, Democracy and Citizenship: The debate over history's role in teaching citizenship and patriotism. Retrieved from www.oah.org/reports/tradhist.html

- Pahl, R. H. (2003a). Assessment traps in K-12 social studies. *The Social Studies*, 94(5), 212 - 215.
- Pahl, R. H. (2003b). Introduction to TSS Special Section: Social Studies Standardized Testing -- Helpful or Harmful? *The Social Studies*, 94(5), 201 - 206.
- Pahl, R. H. (2003c). Social Studies Standardized Testing -- "It's déjà vu all over again!". *The Social Studies*, 94(5), 201 - 206.
- Parker, W. C. (1997). Navigating the unity/diversity tension in education for democracy. *The Social Studies*, 88(1), 12 - 17.
- Pedulla, J. J., Abrams, L. M., Madaus, G. F., Russell, M. K., Ramos, M. A., & Miao, J. (2003). *Perceived effects of state-mandated testing programs on teaching and learning: Findings from a National Survey of Teachers*. Chestnut Hill: Lynch School of Education, Boston College.
- Pinar, W. F. (1996). The curriculum. In J. E. Kincheloe & S. R. Steinberg (Eds.), *Thirteen questions: Reframing education's conversation* (2nd ed., pp. 23 - 30). New York: Peter Lang.
- Popham, W. J. (2001). *The truth about testing: An educator's call to action*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Popkewitz, T. S., & St. Maurice, H. (1991). Social studies education and theory: Science, knowledge and history. In J. P. Shaver (Ed.), *Handbook of Research on Social Studies Teaching and Learning*. New York: Macmillan Publishing Company.
- Porter, A. (2006). Curriculum Assessment. In J. Green, G. Camilli & P. B. Elmore (Eds.), *Handbook of Complementary Methods in Education Research* (pp. 141 - 159). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Porter, A., & Polikoff, M. (2008). The Quality of Content Analyses of State Student Achievement Tests and Content Standards. *Educational Measurement: Issues and Practice*, 27(4), 2 - 14.

- Porter, A., & Smithson, J. (2001). Are content standards being implemented in the classroom? A methodology and some tentative answers. In S. H. Fuhrman (Ed.), *From the capitol to the classroom: Standards-based reform in the states -- One hundredth yearbook of the National Society for the Study of Education, Part II* (pp. 60-80). Chicago: University of Chicago Press.
- Porter, L. (2009). Conversation with Department Chair. In D. Terrell (Ed.). Westford, MA.
- Pryor, C. R. (2006). Pre-service to in-service changes in beliefs: A study of intention to become a democratic practitioner. *Theory and Research in Social Education*, 34(1), 98 - 123.
- Putnam, R. (2000). *Bowling alone: The collapse and revival of American community*. New York: Simon and Schuster.
- Ravitch, D. (2001). *Left back: A century of battles over school reform*. New York: Simon & Schuster.
- Ravitch, D., & Finn, J., C.E. (1987). *What do our 17-year-olds know?: A report on the first National Assessment of History and Literature*. New York: Harper & Row.
- Rees, J. (2003). A crisis over consensus: Standardized testing in American history and student learning. *Radical Pedagogy*.
- Robinson, D., & Groves, J. (2004). *Introducing philosophy*. Cambridge, UK: Totem Books.
- Robinson, D., & Groves, J. (2005). *Introducing Philosophy*. Thriplow, Cambridge: Totem Booksd.
- Rochester, J. M. (2003). The training of idiots. In J. Leming, L. Ellington & K. Porter (Eds.), *Where Did Social Studies Go Wrong*. Washington, D.C.: Thomas B. Fordham Foundation.
- Rorty, R. (1999). *Philosophy and social hope*. New York: Penguin Books.

- Ross, E. W. (1987). Teacher perspective development: A study of preservice social studies teachers. *Theory and Research in Social Education, 15*(4), 225 - 243.
- Ross, E. W. (Ed.). (1997). *The Social Studies Curriculum: Purposes, Problems, and Possibilities*. Albany: State University of New York Press.
- Rothman, R. (2004). *Imperfect matches: The alignment of standards and tests*. Washington, DC: National Research Council.
- Rothstein, R. (2004). We are not ready to assess history performance. *The Journal of American History, 90*(4), 1381 - 1391.
- Salinas, C. (2006). Teaching in a high-stakes testing setting: What becomes of teacher knowledge? In S. G. Grant (Ed.), *Measuring History: Cases of state-level testing across the United States* (pp. 338). Greenwich, CT: Information Age Publishing.
- Salmans, S. (1987). Learning by rote: A new appraisal. *The New York Times*.
- Saunders, R. M. (1996). National standards for United States history: Is the report practical? Would Dewey approve? *The Social Studies, 87*(2), 63 - 67.
- Savage, T. V. (2003). Assessment and quality social studies. *The Social Studies, 94*(5), 201 - 206.
- Schneider, D. (1995). A reaction to "The Storm of Controversy Continues". *The Social Studies, 86*(6), 275.
- Seixas, P. (2000). Schweigen! Die Kinder! Or Does Postmodern History Have a Place in the Schools? In P. S. Stearns, P. & S. S. Wineburg (Eds.), *Knowing, Teaching and Learning History*. New York.
- Shenkman, R. (2008). *Just How Stupid Are We? Facing the **Truth** about the American Voter*. New York: Basic Books.
- Sleeter, C. (1995). Curriculum controversies in multicultural education. In E. Flaxman & H. Passow (Eds.), *Changing populations, changing schools: Ninety-Fourth*

yearbook of the National Society for the Study of Education. Chicago: University of Chicago Press.

- Sleeter, C. (2000). Creating an empowering multicultural curriculum. [Conceptual]. *Race, Gender & Class*, 7(3), 178.
- Sleeter, C., & Delgado Bernal, D. (2004). Critical pedagogy, critical race theory, and antiracist education: Implications for multicultural education. In J. A. Banks & C. A. McGee Banks (Eds.), *Handbook of research on multicultural teaching* (2nd ed., pp. 240 - 258). San Francisco: Josse-Bass.
- Sleeter, C., & Grant, C. (1987). An analysis of multicultural education in the United States. [Conceptual Review]. *Harvard Educational Review*, 57(4), 421 - 444.
- Smiley, M. (1999). Pragmatic inquiry and democratic politics. *American Journal of Political Science*, 43(2), 629 - 647.
- Smith, A. M. (2006). Negotiating control and protecting the private: History teachers and the Virginia Standards of Learning. In S. G. Grant (Ed.), *Measuring History: Cases of state-level testing across the United States* (pp. 338). Greenwich, CT: Information Age Publishing.
- Stanley, W. B. (2005). Social studies and the social order: Transmission or transformation? *Social Education*, 69(5), 282.
- Stanley, W. B., & Nelson, J. R. (Eds.). (1994). *The foundations of social education in historical context*. New York: St. Martin's Press.
- Stern, S. M. (1994). Beyond the rhetoric: An historian's view of the "national" standards for United States history. *Journal of Education*, 173(3), 61 - 71.
- Stodolsky, S. S., & Grossman, P. L. (1995). The impact of subject matter on curricular activity: An analysis of five academic subjects. *American Educational Research Journal*, 32(2), 227 - 249.

- Stotsky, S. (2004). *The stealth curriculum: Manipulating America's history teachers*. Washington, D.C.: The Thomas B. Fordham Foundation.
- Swanson, C. B. (2008). Quality Counts, 2008: State of the states: Grading the states. *Education Week*, 27(18).
- Tabachnick, R., & Zeichner, K. M. (1984). The impact of the student teaching experience on the development of teacher perspectives. *Journal of Teacher Education*, 35(6), 28 - 36.
- Talisse, R. B. (2004). Introduction: Pragmatism and deliberative politics. *Journal of Speculative Philosophy*, 18(1), 1 - 8.
- Thornton, S. J. (1991). Teacher as curricular-instructional gatekeeper in social studies. In J. P. Shaver (Ed.), *Handbook of research on social studies teaching and learning* (pp. 661). New York: McMillan Publishing Company.
- Thornton, S. J. (1994). The social studies near century's end: Reconsidering patterns of curriculum and instruction. *Review of Research in Education*, 20, 223 - 254.
- Thornton, S. J. (2001). From content to subject matter. *The Social Studies*, 92(6), 237 - 242.
- Thornton, S. J. (2008). Continuity and change in social studies curriculum. In L. S. Levstik & C. A. Tyson (Eds.), *Handbook of Research in Social Studies Education* (pp. 402). New York: Routledge: A Taylor and Francis Group.
- Tocqueville, A. d. (2000). *Democracy in America* (H. C. Mansfield & D. Winthrop, Trans.). Chicago: The University of Chicago Press.
- Tyack, D., & Cuban, L. (1997). *Tinkering Toward Utopia: A Century of Public School Reform* (Reprint (March 25, 1997) ed.). Cambridge, Mass: Harvard University Press.

- Vansledright, B., & Grant, S. G. (1994). Citizenship education and the persistent nature of classroom teaching dilemmas. *Theory and Research in Social Education*, 22(3), 305 - 339.
- Vaznis, J. (2009a, January 11, 2009). Officials may postpone charter school openings MCAS history test also targeted amid fiscal woes. *The Boston Globe*, p. B1.
- Vaznis, J. (2009b, February 25, 2009). State delays debut of MCAS history exam by 2 years. *The Boston Globe*.
- Vinson, K. D. (1996). Power/Knowledge and Instructional Approach: A Survey of High School Social Studies Teachers. Unpublished Dissertation. University of Maryland.
- Vinson, K. D. (1998). The "Traditions" Revisited: Instructional Approach and High School Social Studies Teachers. *Theory and Research in Social Education*, 26(1), 50 - 82.
- Vogler, K. (2006). The impact of a high school graduation examination on Mississippi social studies teachers' instructional practices. In S. G. Grant (Ed.), *Measuring History: Cases of state-level testing across the United States* (pp. 338). Greenwich, CT: Information Age Publishing.
- Volger, K. (2003). Where does social studies fit in a high-stakes testing environment? *The Social Studies*, 94(5), 207 - 211.
- Wade, R. (1993). A challenge to the Social Studies: A response to Leming's 'Ideological Perspectives within the Social Studies Profession.' Reaction. *Theory and Research in Social Education*, 21(2), 156-160.
- Webb, N. L. (2002). Depth-of-Knowledge Levels for Four Content Areas. Retrieved December 1, 2008, 2008, from <http://facstaff.wcer.wisc.edu/normw/All%20content%20areas%20%20DOK%20levels%2032802.doc>
- Webb, N. L. (2009). DOK. In D. Terrell (Ed.). Nashua, NH.

- Westheimer, J., & Kahne, J. (2004). What kind of citizen? The politics of educating for democracy. *American Educational Research Journal*, 41(2), 237 - 269.
- White, C. S. (1982). A validation study of the Barth-Shermis social studies preference scale. *Theory and Research in Social Education*, 10(2), 1 - 120.
- Whittington, D. (1991). What have 17-year-olds known in the past? [Empirical]. *American Educational Research Journal*, 28(4), 759 - 780.
- Wideen, M., Mayer-Smith, J., & Moon, B. (1998). A critical analysis of the research on learning to teach: Making the case for an ecological perspective on inquiry. *Review of Educational Research*, 68(2), 130 - 178.
- Will, G. (2008). Ed schools vs. education: Prospective teachers are expected to have the correct "disposition," proof of which is espousing "progressive" political beliefs. *Newsweek*.
- Wilson, E. K., Konopak, B. C., & Readence, J. E. (1994). Preservice teachers in secondary social studies: Examining conceptions and practices. *Theory and Research in Social Education*, 22(3), 364 - 379.
- Wilson, S. M., & Wineburg, S. S. (1988). Peering at history through different lenses: The role of disciplinary perspectives in teaching history. *Teachers College Record*, 89(4), 526 - 539.
- Wineburg, S. S. (2004). Crazy for history. *The Journal of American History*, 90(4), 1401 - 1414.
- Wisconsin Center for Education Research. (2008). K-12 Social Studies Content Taxonomy. Wisconsin Center for Education Research.
- Wronski, S. P. (1993). Persistent issues in the social studies. In V. S. Wilson, J. A. Litle & G. L. Wilson (Eds.), *Teaching Social Studies: Handbook of trends, issues, and implications for the future* (pp. 5 - 23). Westport, CT: Greenwood Press.

- Yeager, E. A., & Pinder, M. (2006). The impact of a high school graduation examination on Mississippi social Studies teachers' instructional practices. In S. G. Grant (Ed.), *Measuring History: Cases of state-level testing across the United States* (pp. 338). Greenwich, CT: Information Age Publishing.
- Yeh, S. (2002). Tests worth teaching to: Constructing state-mandated tests that emphasize critical thinking. *Educational Researcher*, 30(9), 12 - 17.
- Yen, H. (2009, July 20, 2009). Voting rate dips in 2008 as older whites stay at home. *The Associated Press*.
- Yon, M., & Passe, J. (1994). Developing perspectives of social-studies instruction: A longitudinal case study. *Journal of Research and Development in Education* 28(1), 31 - 42.
- Zeichner, K. M., & Liston, D. P. (1987). Teaching student teachers to reflect. *Harvard Educational Review*, 57(1), 23-48.
- Zumwalt, K., & Craig, E. (2005). Teachers' characteristics: Research on the demographic profile. In M. Cochran-Smith & K. M. Zeichner (Eds.), *Studying teacher education: The report of the AERA panel of research and teacher education* (pp. 111 - 156). Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

APPENDIX A: WORK PLAN

Task	Date
Power analysis using data on MassDESE website	November 2008
Defend dissertation proposal	December 2008
Conduct pilot studies and revise survey instrument	December 2008
Secure Institutional Review Board (IRB) approval for study	January 2008
Identification of teachers and department chairs via school websites	January 2008
Create a database of the public high schools in Massachusetts.	January 2008
Randomly number each of the schools.	
Survey Instrument:	
1. Creation of instrument that conforms to IRB standards.	November 2008
2. Review existing Vinson instrument establish alignment between the inferences and conclusions Vinson was able to draw with his study and the inferences and conclusions I hope to reach.	November 2008
3. Reliability: to determine reliability of Vinson instrument, establish expert panel and have expert panel sort items in instrument into constructs.	December 2008
4. Pilot: Pilot survey to existing group of pre-service history and social studies teachers for errors in wording, double barreled questions etc.	December 2008

5. Revise: Revise instrument according to results from panel and pilot study	January 2009
Survey Administration: (Using <i>Dillman's Tailored Design Method</i>)	
• Send notice of survey administration postcard	February 2009
• 1 st administration: Send survey with cover letter, token gift and options for viewing results of the survey	February 2009
• Telephone and/or email reminder to complete survey	February, 2009
• 1 st administration due date for electronic and paper version	March 1, 2009
• Check sample against returned surveys, prepare for second administration	March 2009
• 2 nd administration: Send survey w/cover letter, token gift and options for viewing results of the survey	March 2009
• Telephone and/or email reminder	March, 2009
• End survey administration	April 6, 2009
Data Input	April 2009
Data Analysis	Spring 2009

APPENDIX B: ORIGINAL VINSON SURVEY INSTRUMENT

HIGH SCHOOL SOCIAL STUDIES TEACHER QUESTIONNAIRE

Part I. Directions: Please read and answer each of the following questions. Please circle the one numbered response that most closely represents your belief with respect to each item.

1. High school social studies teachers should utilize the methods of critical thinking so that their students can see how in the United States the powerful often disregard the rights of the powerless.

Strongly Disagree 1 2 3 4 5 Strongly Agree

2. High school social studies content should emphasize the histories and cultures of women, people of color, and members of lower socio-economic classes.

Strongly Disagree 1 2 3 4 5 Strongly Agree

3. High school social studies teachers should use methods in which students are allowed to identify and solve their own problems.

Strongly Disagree 1 2 3 4 5 Strongly Agree

4. Lecture should be used as an instructional strategy by high school social studies teachers because it enables students to acquire and retain a significant number of specific facts and concepts.

Strongly Disagree 1 2 3 4 5 Strongly Agree

5. One purpose of high school social studies education should be providing students the opportunity to be critical of America's historical and contemporary institutions and traditions with respect to cultural, economic, social, and political injustice and inequality.

Strongly Disagree 1 2 3 4 5 Strongly Agree

6. High school social studies teachers should allow students to identify their own problems, to develop their own hypotheses, to collect their own data, and to draw their own conclusions--even if they disagree with the conclusions of teachers and/or with the knowledge and values of a majority of the members of American society.

Strongly Disagree 1 2 3 4 5 Strongly Agree

7. One purpose of high school social studies education should be the strengthening of American democracy not by instilling in students a common body of information but by encouraging them to recognize and solve their own individual problems.

Strongly Disagree 1 2 3 4 5 Strongly Agree

8. High school social studies students should learn and practice the process of conducting historical research in exactly the same way as professional historians do.

Strongly Disagree 1 2 3 4 5 Strongly Agree

9. When presented a problem to solve, high school social studies students should be able to identify it as characteristic of a specific social science discipline and to solve it according to the processes representative of that specific social science discipline.

Strongly Disagree 1 2 3 4 5 Strongly Agree

10. High school social studies content should be standard throughout a school district and should be developed by a district curriculum committee.

Strongly Disagree 1 2 3 4 5 Strongly Agree

11. High school social studies content should be teacher-selected based upon maximizing student success and growth.

Strongly Disagree 1 2 3 4 5 Strongly Agree

12. High school social studies teachers should utilize methods based upon students' cultural backgrounds.

Strongly Disagree 1 2 3 4 5 Strongly Agree

13. High school social studies teachers should emphasize teaching students to solve problems that have been identified as important by professional social scientists.

Strongly Disagree 1 2 3 4 5 Strongly Agree

14. High school social studies teachers should utilize cooperative learning and other methods that increase students' levels of self-esteem.

Strongly Disagree 1 2 3 4 5 Strongly Agree

15. High school social studies content should be based upon student-identified problems.

Strongly Disagree 1 2 3 4 5 Strongly Agree

16. High school social studies content should focus upon past and contemporary examples of racism, sexism, and elitism and how to reduce/eliminate them in the future.

Strongly Disagree 1 2 3 4 5 Strongly Agree

17. High school social studies content should be determined by professional social scientists (e.g., historians, geographers, political scientists, etc.).

Strongly Disagree 1 2 3 4 5 Strongly Agree

18. One purpose of high school social studies should be instructing students how to work for racial, gender, and economic justice and equality.

Strongly Disagree 1 2 3 4 5 Strongly Agree

19. One purpose of high school social studies education should be developing in students a strong and positive sense of self-esteem.

Strongly Disagree 1 2 3 4 5 Strongly Agree

20. The "social studies" should be synonymous with the "social sciences."

Strongly Disagree 1 2 3 4 5 Strongly Agree

21. High school social studies teachers should utilize as many instructional strategies as possible in order to maximize the number of students who are successful.

Strongly Disagree 1 2 3 4 5 Strongly Agree

22. One purpose of high school social studies education should be ensuring the academic success of all students.

Strongly Disagree 1 2 3 4 5 Strongly Agree

23. The content of each high school social studies course should focus on a single, specific social science discipline (e.g., grade 10, world history, grade 11, US history, etc.).

Strongly Disagree 1 2 3 4 5 Strongly Agree

24. High school social studies content should be interdisciplinary.

Strongly Disagree 1 2 3 4 5 Strongly Agree

25. Student assessment in high school social studies should be based primarily upon objective (e.g., multiple choice) tests that closely parallel actual course content.

Strongly Disagree 1 2 3 4 5 Strongly Agree

26. High school social studies content should be student-selected based upon student perceptions of their own aptitudes and interests.

Strongly Disagree 1 2 3 4 5 Strongly Agree

27. One purpose of high school social studies education should be developing within students the ability to solve problems that are relevant to their own specific, real life, individual, and social experiences.

Strongly Disagree 1 2 3 4 5 Strongly Agree

28. In general, high school social studies textbooks are effective in presenting a body of information that is both important and appropriate for all students.

Strongly Disagree 1 2 3 4 5 Strongly Agree

29. One purpose of high school social studies education should be the development within students of the belief that democracy is the best form of government.

Strongly Disagree 1 2 3 4 5 Strongly Agree

30. One purpose of high school social studies education should be the transmission of traditional American history, geography, culture, and values.

Strongly Disagree 1 2 3 4 5 Strongly Agree

Part II. Directions: Please read and answer each of the following questions. Please circle the one numbered response that most closely represents your belief with respect to each item.

31. High school social studies education should transmit to all students a common body of traditional American knowledge, knowledge based upon mainstream history, culture, and values, using methods such as textbook readings, lectures, and objective examinations.

Strongly Disagree 1 2 3 4 5 Strongly Agree

32. High school social studies education should teach the concepts and methods of the individual social sciences (e.g., culture and ethnography from anthropology, inflation and cost-benefit analysis from economics) as a way of solving particular types of problems falling within the domain of the social sciences; social studies teaching and learning should consist of teachers and students acting as if they were social scientists (e.g., anthropologists, economists).

Strongly Disagree 1 2 3 4 5 Strongly Agree

33. High school social studies education should teach students to learn to solve problems and to make decisions; it should focus on individual and social problems perceived by students as relevant to their real life experiences; they should learn to solve such problems by way of selecting and testing their own hypotheses, collecting and analyzing their own data, using whatever resources are necessary, and they should strive and be encouraged to become independent problem solvers and decision makers.

Strongly Disagree 1 2 3 4 5 Strongly Agree

34. High school social studies education should critically examine America's institutions--both historical and contemporary--in terms of social injustice and/or social inequality; teachers and students should identify and investigate, for example, instances of cultural, economic, ethnic, gender, and racial discrimination; teachers should use a variety of instructional strategies, emphasizing how some Americans have been and are still able to exploit others.

Strongly Disagree 1 2 3 4 5 Strongly Agree

35. High school social studies education should instill in students a sense of self-worth, a sense of self-esteem, and a sense of self-efficacy; they should study examples of how individuals or groups of people have been able to overcome some difficulty or difficulties in order to achieve success; teachers should utilize techniques which allow all of their students to learn, for example cooperative learning and individualized instruction.

Strongly Disagree 1 2 3 4 5 Strongly Agree

With which of the statements above do you most strongly agree? (Please circle one [1].)

#31 #32 #33 #34 #35

APPENDIX C: REVISED SURVEY INSTRUMENT USED IN THIS STUDY



Random ID:
«School_ID»

Massachusetts History and Social Studies Teacher Survey

Please check the box to the left to indicate that you have received the cover letter “An Invitation to Participate in Voluntary Research”

If you would like access to an electronic executive summary of the data gathered from this survey, please indicate your interest by providing your email address in the space below. Your email address will not be sold, used for identification purposes or be used in connection with your answers to this survey in any way.

Participant’s Email Address: _____

Directions: Please answer the following questions on your beliefs about standards and assessments for high school history and social studies. Please indicate the strength of your agreement or disagreement with each of the following statements by marking the corresponding box with an “x”.

Standards and Assessments	Strongly Agree	Agree	Disagree	Strongly Disagree
1. The MCAS 10/11 th grade mandated test in US History is compatible with my daily instruction.				
2. My district’s curriculum is aligned with the state mandated curriculum.				
3. Scores on the state-wide mandated test accurately reflect the quality of education students have received.				
4. If I teach to the state standards or frameworks, students will do well on the state-mandated test.				
5. My tests are in the same format as the state mandated test.				
6. The state mandated testing program leads some teachers in my school to teach in ways that contradict their own ideas of good education practice.				
Please write any comments about the state assessments here including changes you would like to see or things you would like to see retained for future administrations of the MCAS high school US History test. You may attach additional sheet(s) if you need more space.				

Directions: Please answer the following questions on your beliefs about the purpose of high school history and social studies. Please indicate the strength of your agreement or disagreement with each of the following statements by marking the corresponding box with an “x”.

As a result of high school social studies, students should ...	Strongly Agree	Agree	Disagree	Strongly Disagree
7. be able to methodically analyze social problems.				
8. view American institutions and traditions with a critical eye.				
9. be able to demonstrate knowledge of traditional American history, geography, culture, and values.				
10. identify and work through problems according to the processes used by social scientists.				
11. work toward social justice and equality.				
12. possess positive self-esteem.				
13. defend the methods of social science investigation, even when the results obtained contradict popular opinion.				
14. demonstrate increased understanding of how they can succeed personally and academically.				
15. methodically inquire into problems that are relevant to their own individual and social experiences.				
16. believe that democracy is the best form of government.				



Random ID:
«School_ID»

Massachusetts History and Social Studies Teacher Survey

Directions: Please answer the following questions on the pedagogical choices of high school history and social studies teachers. Please indicate the strength of your agreement or disagreement with each of the following statements by marking the corresponding box with an "x".

High school social studies teachers should...	Strongly Agree	Agree	Disagree	Strongly Disagree
17. plan lessons that allow time for students to discuss power inequity in the United States.				
18. plan lessons that allow students to methodically inquire into their own problems.				
19. lecture, because it enables students to acquire and retain significant facts and concepts.				
20. teach students the methods of inquiry (i.e. how to identify their own problems, develop hypotheses, collect data, and draw their own conclusions) as a means to students' personal fulfillment.				
21. teach students the research skills that professional social scientists use.				
22. plan lessons that ask students to practice skills of democratic activism (e.g. discussion, debate).				
23. emphasize teaching students to solve problems that have been identified as important by professional social scientists.				
24. utilize methods that increase students' levels of self-esteem (e.g. differentiated instruction).				
25. utilize as many instructional strategies as possible in order to maximize the number of students who are successful.				
26. assess students primarily with objective (e.g., multiple choice) tests that parallel course content.				

Directions: Please answer the following questions on your beliefs about high school history and social studies content. Please indicate the strength of your agreement or disagreement with each of the following statements by marking the corresponding box with an "x".

High school social studies content should ...	Strongly Agree	Agree	Disagree	Strongly Disagree
27. emphasize the histories of minority groups (e.g. women, people of color, & the working class).				
28. be standardized for school districts.				
29. be selected based upon maximizing student success and growth.				
30. be determined by professional social scientists (e.g., historians, geographers, political scientists, etc.).				
31. be based upon student-identified problems.				
32. be based on the research procedures and findings from social science authorities.				
33. arise from investigating problems derived from students' own needs and interests.				
34. focus on past and contemporary examples of racism, sexism, and elitism and how to reduce/eliminate them in the future.				
35. be student-selected based upon student perceptions of their own aptitudes.				
36. come from textbooks, because they contain information that is important and appropriate.				



Random ID:
«School_ID»

Massachusetts History and Social Studies Teacher Survey

Directions: Please indicate the strength of your agreement or disagreement with each of the following statements by marking the corresponding box with an "x".

	Strongly Agree	Agree	Disagree	Strongly Disagree
37. High school social studies education should transmit to students a common body of traditional American knowledge based on mainstream history, culture, and values, using methods such as textbook readings, lectures, and objective examinations.				
38. High school social studies education should teach the concepts and methods of the social sciences as a way of solving problems identified by social scientists; social studies teaching and learning should consist of teachers and students acting as social scientists (e.g., anthropologists, economists).				
39. High school social studies education should teach students to consider problems and to make decisions; it should focus on individual and social problems perceived by students as relevant to their life experiences; social studies teaching and learning should consist of students selecting and testing their own hypotheses, collecting and analyzing their own data, and making decisions based on the data.				
40. High school social studies education should critically examine America's institutions—both historical and contemporary—in terms of social injustice and/or social inequality; teachers should use a variety of instructional strategies, emphasizing how some Americans have been and are still able to exploit others; social studies is a means to educate and guide active democratic citizens.				
41. High school social studies education should instill in students a sense of self-worth, a sense of self-esteem, and a sense of self-efficacy; they should study examples of how individuals or groups of people have been able to overcome difficulties to achieve success; teachers should utilize techniques which allow all of their students to learn (e.g. cooperative learning and individualized instruction).				

42. With which of the statements above do you most strongly agree? (Please circle one [1].)

Statement #37 Statement #38 Statement #39 Statement #40 Statement #41

Directions: Please mark all that apply with an "x".

43. What subject(s) are you teaching in the academic year 2008 - 2009? Please mark all that apply.

_____ United States History

_____ World History

_____ European History

_____ Other (Please indicate in space provided to the right.) _____

44. What grade level(s) are you teaching in the academic year 2008-2009? Please mark all that apply.

_____ Ninth

_____ Tenth

_____ Eleventh

_____ Twelfth

OTHER: _____

~PLEASE TURN TO LAST PAGE~



Random ID:
«School_ID»

Massachusetts History and Social Studies Teacher Survey

Directions: Please mark the appropriate response with an "x."

45. How do you prepare your students for your state-mandated test? Mark ALL that apply.

- | | |
|---|--|
| <input type="checkbox"/> I do no special test preparation. | <input type="checkbox"/> I teach the standards of frameworks known to be on the test. |
| <input type="checkbox"/> I teach test-taking skills. | <input type="checkbox"/> I provide students with items similar to those on the test. |
| <input type="checkbox"/> I encourage students to work hard and prepare. | <input type="checkbox"/> I provide test-specific preparation materials developed commercially or by the state. |
| <input type="checkbox"/> I provide rewards for test completion. | <input type="checkbox"/> I provide students with released items from the state-mandated test. |

46. How many years of teaching experience do you have including this year?

- 1 to 3 years.
 4 to 8 years.
 9 to 12 years.
 13 to 20 years.
 Over 20 years

47. What is your gender?

- Female
 Male

48. Please mark the appropriate range for your age.

- 20-30
 31-40
 41-50
 51-60
 61+

49. Mark all of the following categories that best describe you.

- African American
 American Indian or Alaskan Native
 Asian American
 White/Caucasian American
 Pacific Islander
 Hispanic
 Other, please specify: _____

50. Which of the following best describes you?

- Very Liberal
 Liberal
 Moderate
 Conservative
 Very Conservative

Please be sure that you checked the box at the top of the first page that indicates you received the "Invitation to Participate in Voluntary Research"

Thank you for your time

Questions 1 – 6 and 43 – 50 adapted from Pedulla, J. J., Abrams, L. M., Madaus, G. F., Russell, M. K., Ramos, M. A., & Miao, J. (2003). Perceived effects of state-mandated testing programs on teaching and learning: Findings from a National Survey of Teachers. Chestnut Hill: Lynch School of Education, Boston College.

Questions 7-42 adapted from Vinson, K. D. (1996). Power/Knowledge and Instructional Approach: A Survey of High School Social Studies Teachers (pp. 250): University of Maryland.

APPENDIX D: INVITATION TO PARTICIPATE IN VOLUNTARY RESEARCH



INVITATION TO PARTICIPATE IN VOLUNTARY RESEARCH

Dear <<First Name>> <<Last Name>>,

My name is Dianna Terrell and I am a doctoral candidate in the Lynch School of Education at Boston College. I am conducting an anonymous survey on teachers' opinions about the 10th/11th grade MCAS test in US History and how it connects to teaching history and social studies. Many people make assumptions about teachers' beliefs and how they translate into pupil's test scores on high stakes tests. However, little actual research has been done to explore that relationship.

Your feedback could not come at a more crucial juncture. Even as I write, the Massachusetts Commissioner of Education plans to request a suspension of this year's pilot administration of the test. Some teachers see this as a welcome opportunity to address what they see as fundamental problems with the U.S. History test. Still, other teachers are deeply concerned that this could mean history and social studies will be relegated to the back burner. Hearing what social studies and history teachers in Massachusetts think about this move is absolutely critical.

As with most research, there may be unknown risks. To participate, you must be 18 years or older. The survey is voluntary and may be discontinued at any time. Withdrawal from participation will not result in denial of entitled benefits. Aggregate findings from the study will be shared with my dissertation committee and with state officials who are developing the MCAS US History test. The results will be reported for the group of respondents as a whole. Random IDs have been assigned to participants only to track participation in the survey but will not be used to link your anonymous survey responses to your name. I assure you that individual answers to survey questions will be kept confidential. If you are interested in seeing the results of this survey, please indicate your interest by providing your email address at the top of the first page of the survey.

As a former high school history teacher, I can recall the many things on the faculty's agenda. I sincerely hope you find interest in this study and can allocate time to lend your voice to this research. The survey will take approximately 10 minutes. Please answer the questions to your comfort level. Since your answers are to remain anonymous, PLEASE DO NOT PUT YOUR NAME ON THIS SURVEY.

Please return this survey in the self addressed stamped envelope provided by March 5, 2009. Participants who return completed surveys on or before this date will be entered into a raffle to win one of three \$20 dollar gift cards to Barnes and Noble Bookstores. You are also welcome to access the electronic version of the survey with your seven-digit random ID at

**www2.bc.edu/~terreldi
RANDOM ID: 8372664**

If you have any questions about the authenticity of this study or my credentials, please contact my dissertation chair, Dr. Marilyn Cochran-Smith at Marilyn.Cochran-Smith.1@bc.edu or via her website at <http://www.bc.edu/schools/lsoe/facultystaff/faculty/cochran-smith.html>.

Thank you for your consideration.

Sincerely,

A handwritten signature in black ink that reads "Dianna Terrell". The signature is written in a cursive style with a large, stylized "D" and "T".

APPENDIX E: CONTENT ANALYSIS MATRIX

Depth of Knowledge				Content Orientation	
Level One	Level Two	Level Three	Level Four		
<p>Level One Recall, recognize, identify, list, define, recite, and reproduce information. Describe and explain information in maps, charts, tables graphs or drawings.</p>	<p>Level Two Basic reasoning (may not know answer could but can use reason to determine answer), contrast and compare, convert, classify or sort. Descriptions & explanations result in understanding how & why.</p>	<p>Level Three Complex reasoning, justify how & why. Apply using evidence. Use concepts to solve problems. Propose & evaluate solutions. Recognize & explain misconceptions. Make connections across time & place.</p>	<p>Level Four Extended reasoning. Plan, investigate or develop something over time. Connect & relate ideas & concepts within & among content areas. Use evidence from multiple sources. Examine and explain alternative perspectives and solutions. Make predictions.</p>	<p>RI for Social Ends Content ...emphasizes minority/multicultural groups, racism, sexism, power, classism, inequality, authority, institutions and how to bring about social justice and equality.</p>	<p>RI for Personal Ends Content ...emphasizes building relationships between people. Developing self reflection, empathy, human relations (where do you stand on...)</p>
				<p>Reflective Inquiry Content ...focuses on social problems, controversial issues, critique of authority, broad range of political content, increases students political tolerance and open mindedness.</p>	<p>Social Science Discipline Content ...focuses on content and problems in the social sciences (i.e. economics, sociology, psychology, sociology, geography). Predetermined answers.</p>
				<p>Cultural Transmission Content ...focuses on discrete factual, traditional and/or canonical knowledge about history and government including important American heroes, symbols and dates.</p>	
<p>Rater ID and Feedback</p>					

APPENDIX F: SAMPLE QUESTIONS FROM THE MCAS-US TEST

Sample Items from the 10th/11th Grade U.S. History MCAS Test

- 1 In his *Farewell Address*, President Washington warned the United States to avoid which of the following?
- A. trade relationships with European nations
 - B. diplomatic relations with European powers
 - ✓ C. permanent alliances with European powers
 - D. immigration agreements with European nations
- 2 How did the cotton gin affect the Southern economy from 1800 to 1860?
- A. It encouraged industrialization in the South.
 - B. It promoted economic equality in the South.
 - ✓ C. It strengthened Southerners' reliance on slavery.
 - D. It increased Southerners' use of indentured servants.
- 3 Which of the following was an effect of the Taft-Hartley Act (1947) on labor unions?
- A. It prohibited labor unions from being segregated.
 - ✓ B. It restricted the organizing activities of labor unions.
 - C. It allowed labor unions to use dues for political activity.
 - D. It guaranteed health benefits for members of labor unions.
- 4 Which of the following is a legal requirement of all United States citizens?
- A. to provide aid to the poor
 - B. to vote in federal elections
 - ✓ C. to serve on a jury if called
 - D. to work for the community
- 5 The Battle of Vicksburg was significant because it
- A. marked the end of Confederate invasions into Union territory.
 - B. destroyed the South's greatest city, devastating Confederate morale.
 - ✓ C. gave the Union control of the Mississippi and split the Confederacy in two.
 - D. created a moment appropriate for President Lincoln to proclaim Emancipation.

Sample Items from the 10th/11th Grade U.S. History MCAS Test

- 6** How did World War II impact the lives of African Americans from 1941–1945?
- A. African Americans were widely elected to public office.
 - B. African Americans served with whites in desegregated military units.
 - C. Civil rights reforms greatly improved the social status of African Americans.
 - ✓ D. Defense industry jobs led many African Americans to migrate from the South.
- 7** During the Great Depression, the United States economy was **mainly** characterized by
- A. inflation.
 - ✓ B. deflation.
 - C. increasing wages.
 - D. over-consumption.

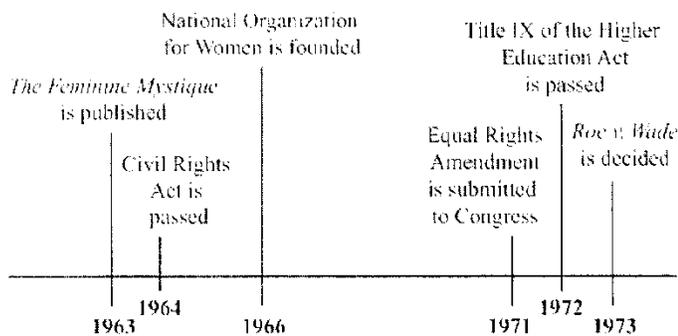
Sample Items from the 10th/11th Grade U.S. History MCAS Test

Grades 10 and 11 U.S. History: Sample Module

The section below deals with this theme of American history: the evolution of the concepts of personal freedom and the respect for human dignity. The timeline below shows events that are important to this theme.

First, you will answer two multiple-choice questions about the post-World War II women's rights movement. Then you will answer two parts of an open-response question about the effectiveness of the women's rights movement.

Important Events in the United States Women's Rights Movement



- 8 Who wrote *The Feminine Mystique*, a book that helped spark the women's rights movement?
- ✓ A. Betty Friedan
 - B. Phyllis Schlafly
 - C. Eleanor Roosevelt
 - D. Sandra Day O'Connor
- 9 Which of the following was the primary issue that the Equal Rights Amendment attempted to address?
- A. voting rights
 - B. property rights
 - ✓ C. employment rights
 - D. reproductive rights

Sample Items from the 10th/11th Grade U.S. History MCAS Test

Grades 10 and 11 U.S. History: Sample Open-Response Question

- 10 In the 1960s, the women's rights movement regained some of the vigor and enthusiasm that it had lost after the passage of the Nineteenth Amendment, which gave women the right to vote in 1920.
- Identify **one** issue that was important to the women's rights movement in the 1960s and 1970s. Explain why the issue you identified became important to women at this time. Support your answer with information from the timeline and your knowledge of U.S. history.
 - Explain what gains, if any, the women's rights movement has made since the 1960s in resolving the issue you identified in part (a). Support your answer with information from the timeline and your knowledge of U.S. history.

Sample Items from the 10th/11th Grade U.S. History MCAS Test

Grades 10 and 11 U.S. History: Sample Module

In the section below you will be using selected amendments from the Bill of Rights of the United States Constitution.

First, you will answer two multiple-choice questions about the Bill of Rights. Then you will answer two parts of an open-response question about the importance of the Bill of Rights.

Selected Amendments from the Bill of Rights (1791)

Amendment I

Congress shall make no law respecting an establishment of religion, or prohibiting the free exercise thereof; or abridging the freedom of speech, or of the press; or the right of the people peaceably to assemble, and to petition the Government for a redress of grievances.

Amendment II

A well-regulated Militia, being necessary to the security of a free State, the right of the people to keep and bear Arms, shall not be infringed.

Amendment III

No Soldier shall, in time of peace be quartered in any house, without the consent of the Owner, nor in time of war, but in a manner to be prescribed by law.

Amendment IV

The right of the people to be secure in their persons, houses, papers, and effects, against unreasonable searches and seizures, shall not be violated, and no Warrants shall issue, but upon probable cause, supported by Oath or affirmation, and particularly describing the place to be searched, and the persons or things to be seized.

Amendment V

No person shall be held to answer for a capital, or otherwise infamous crime, unless on a presentment or indictment of a Grand Jury, except in cases arising in the land or naval forces, or in the Militia, when in actual service in time of War or public danger; nor shall any person be subject for the same offence to be twice put in jeopardy of life or limb; nor shall be compelled in any criminal case to be a witness against himself, nor be deprived of life, liberty, or property, without due process of law; nor shall private property be taken for public use, without just compensation.

Amendment VI

In all criminal prosecutions, the accused shall enjoy the right to a speedy and public trial, by an impartial jury of the State and district wherein the crime shall have been committed, which district shall have been previously ascertained by law, and to be informed of the nature and cause of the accusation; to be confronted with the witnesses against him; to have compulsory process for obtaining witnesses in his favor, and to have the Assistance of Counsel for his defense.

Amendment VIII

Excessive bail shall not be required, nor excessive fines imposed, nor cruel and unusual punishments inflicted.

Amendment X

The powers not delegated to the United States by the Constitution, nor prohibited by it to the States, are reserved to the States respectively, or to the people.

Sample Items from the 10th/11th Grade U.S. History MCAS Test

Grades 10 and 11 U.S. History: Sample Module Multiple-Choice Questions

- 11 Who among the following **most** influenced the passage of the Bill of Rights?
- A. John Adams
 - B. Benedict Arnold
 - ✓ C. James Madison
 - D. Roger Williams
- 12 One legal protection guaranteed by the Sixth Amendment is that a person accused of a crime must be
- A. given a fair punishment.
 - ✓ B. allowed access to a lawyer.
 - C. tried for the crime only once.
 - D. given a warrant for his arrest.

Grades 10 and 11 U.S. History: Sample Module Open-Response Question

- 13 On September 12, 1787, George Mason of Virginia asked the delegates to the Constitutional Convention to add a bill of rights to the new plan of government. They voted not to do so. Four years later, the first ten amendments were added to the Constitution as the Bill of Rights.
- a. Choose **one** amendment from the selected amendments from the Bill of Rights. Explain why most Americans in 1791 believed the right or rights outlined in the amendment you chose were important enough to make them a part of the Constitution. Support your answer with information from the Bill of Rights and your knowledge of U.S. history.
 - b. Choose **another** amendment from the selected amendments from the Bill of Rights. Explain why most Americans in 1791 believed the right or rights outlined in the amendment you chose were important enough to make them a part of the Constitution. Support your answer with information from the Bill of Rights and your knowledge of U.S. history.

Sample Items from the 10th/11th Grade U.S. History MCAS Test

- 14** Why was the Bill of Rights added to the United States Constitution?
- A. to ensure rights of foreigners
 - B. to ensure slaves' right to vote
 - C. to protect the federal government from the states
 - D. to protect the individual rights of citizens from government abuse

- 15** When people purchase shares of stock in the stock market, they are investing in
- A. corporations.
 - B. labor unions.
 - C. governments.
 - D. political parties.

- 16** The excerpt below is from Frederick Douglass's "Independence Day" speech in 1852.

The rich inheritance of justice, liberty, prosperity and independence, bequeathed by your fathers is shared by you, not by me. The sunlight that brought life and healing to you has brought stripes and death to me. This Fourth of July is yours, not mine.

—Frederick Douglass,
"Independence Day" speech (1852)

In this speech, Frederick Douglass was speaking on behalf of which group?

- A. exiled American Indians
- B. deported American Jews
- C. oppressed Irish Americans
- D. enslaved African Americans

Sample Items from the 10th/11th Grade U.S. History MCAS Test

- 17** The table below shows the changes in the percentage of households owning specific items between 1940 and 1955.

Ownership of Automobiles and Household Appliances

Items Owned	Percentage of U.S. Households Owning Items	
	1940	1955
an automobile	50%	71%
a television	0%	76%
a refrigerator	44%	94%
a washing machine	56%	84%
a clothes dryer	0%	9%
a vacuum cleaner	38%	64%

Paul Boyer, *The American Nation* (1998)

According to the table, which of the following is one way that American households changed between 1940 and 1955?

- A. They had fewer children.
- B. They traveled less often.
- C. They enjoyed increased prosperity.
- D. They purchased fewer consumer goods.

- 18** Which of the following **best** explains why the United States Senate rejected American participation in the League of Nations after World War I?

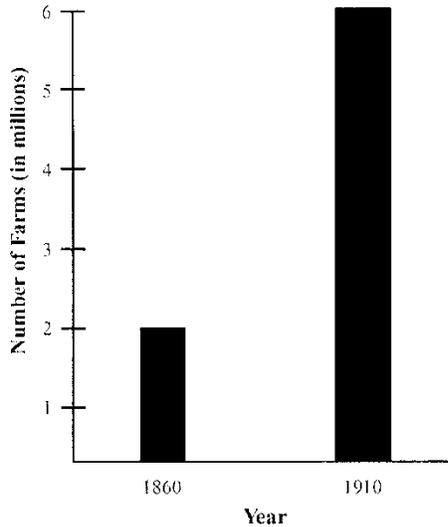
- A. Senators thought league membership would cost too much money.
- B. Senators thought the league would interfere in Latin American affairs.
- C. Senators thought the league would require its members to reduce tariffs.
- D. Senators thought league membership would undermine American sovereignty.

- 19** Which of the following is the **most** important responsibility of the Federal Reserve System?

- A. to set the minimum wage
- B. to regulate gasoline prices
- C. to regulate stock exchanges
- D. to control the money supply

Sample Items from the 10th/11th Grade U.S. History MCAS Test

- 20** The graph below shows the number of farms in the United States in 1860 and 1910.



Paul Johnson, *A History of the American People*
(Harper Perennial: 1997)

How did the Homestead Act of 1862 contribute to the changes shown on the graph?

- A. The government provided land to settlers in the West.
- B. The government gave funds to railroad companies in the West.
- C. The government funded large-scale irrigation projects in the West.
- D. The government promised to buy cash crops from farmers in the West.

- 21** What was the most important result of the Hayes-Tilden agreement (Compromise of 1877) following the presidential election of 1876?

- A. the end of Reconstruction
- B. the passage of the Dawes Act
- C. the rise of multiple radical political parties
- D. the construction of the transcontinental railroad

- 22** The Supreme Court case *Marbury v. Madison* (1803) established which constitutional principle?

- A. due process
- B. judicial review
- C. implied powers
- D. equal protection

- 23** An oligarchy can **best** be described as a government that is ruled by

- A. a king.
- B. the people.
- C. a small elite.
- D. religious leaders.

Sample Items from the 10th/11th Grade U.S. History MCAS Test

In this section, you will consider Abraham Lincoln's Gettysburg Address.

First, you will answer three multiple-choice questions. Then you will answer two parts of an open-response question about the Gettysburg Address.

The Gettysburg Address (1863)

Four score and seven years ago our fathers brought forth on this continent, a new nation, conceived in Liberty, and dedicated to the proposition that all men are created equal.

Now we are engaged in a great civil war, testing whether that nation, or any nation so conceived and so dedicated, can long endure. We are met on a great battle-field of that war. We have come to dedicate a portion of that field, as a final resting place for those who here gave their lives that that nation might live. It is altogether fitting and proper that we should do this.

But, in a larger sense, we can not dedicate—we can not consecrate—we can not hallow—this ground. The brave men, living and dead, who struggled here, have consecrated it, far above our poor power to add or detract. The world will little note, nor long remember what we say here, but it can never forget what they did here. It is for us the living, rather, to be dedicated here to the unfinished work which they who fought here have thus far so nobly advanced. It is rather for us to be here dedicated to the great task remaining before us—that from these honored dead we take increased devotion to that cause for which they gave the last full measure of devotion—that we here highly resolve that these dead shall not have died in vain—that this nation, under God, shall have a new birth of freedom—and that government of the people, by the people, for the people, shall not perish from the earth.

Sample Items from the 10th/11th Grade U.S. History MCAS Test

- 24** Why did President Lincoln travel to Gettysburg to deliver an address in 1863?
- A. to establish headquarters for the Union army
 - B. to dedicate a cemetery for deceased soldiers
 - C. to honor families of confederate volunteers
 - D. to raise money for the abolitionist cause
- 25** The Battle of Gettysburg was an important event in the Civil War **mainly** for which of the following reasons?
- A. Great Britain refused to support the Confederacy after the Union victory.
 - B. The Union victory was the first time the North used African American soldiers in combat.
 - C. The Confederacy never attempted another major invasion of the North after the Union victory.
 - D. Congress immediately ratified the Thirteenth Amendment upon hearing of the Union victory.
- 26** To which of the following events was President Lincoln referring when he stated, "Four score and seven years ago our fathers brought forth on this continent, a new nation . . ."?
- A. the Boston Tea Party
 - B. the Battle of Yorktown
 - C. the signing of the Declaration of Independence
 - D. the ratifying of the Constitution of the United States

Sample Items from the 10th/11th Grade U.S. History MCAS Test

Write your answer to open-response question 14 in the space provided in your Student Answer Booklet.

- 27** In the Gettysburg Address, President Lincoln refers to the nation as having “a new birth of freedom.”
- a. Explain what President Lincoln meant when he called for “a new birth of freedom.” You may support your answer with information from the Gettysburg Address and your knowledge of American history.
 - b. Explain why Lincoln believed a unified nation should be preserved. You may support your answer with information from the Gettysburg Address and your knowledge of American history.

Sample Items from the 10th/11th Grade U.S. History MCAS Test

Mark your answers to questions 15 and 16 in the spaces provided in your Student Answer Booklet.

- 28** The excerpt below is from Richard Nixon's "Silent Majority" speech.

Let historians not record that when America was the most powerful nation in the world we passed on the other side of the road and allowed the last hopes for peace and freedom of millions of people to be suffocated by the forces of totalitarianism.

And so tonight—to you, the great silent majority of my fellow Americans—I ask for your support.

—President Richard Nixon,
"Silent Majority" speech (November 3, 1969)

In his speech, President Nixon asked the American people to support his position on which of the following issues?

- A. the Vietnam War
- B. the War Powers Act
- C. the Watergate scandal
- D. the Middle East peace process

- 29** The quotation below is from a speech given by Senator Albert Beveridge in 1898.

Hawaii is ours; [Puerto] Rico is to be ours; at the prayer of her people Cuba finally will be ours; in the islands of the East . . . the flag of a liberal government is to float over the Philippines . . . The Opposition tells us that we ought not to govern a people without their consent. I answer the rule . . . that all just government derives its authority from the consent of the governed, applies only to those who are capable of self-government.

—Senator Albert Beveridge,
"The March of the Flag" speech (1898)

Which policy was Senator Beveridge advocating in this statement?

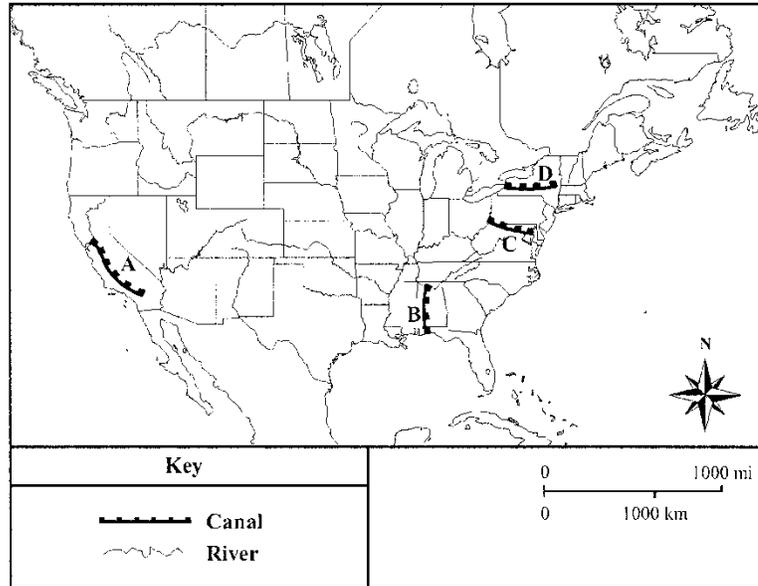
- A. containment
- B. disarmament
- C. imperialism
- D. isolationism

Sample Items from the 10th/11th Grade U.S. History MCAS Test

DIRECTIONS

This session contains seven multiple-choice questions. Mark your answers to these questions in the spaces provided in your Student Answer Booklet.

- 30** The map below shows the continental United States with four canals labeled A, B, C, and D.



Which letter marks the location of the Erie Canal?

- A. A
- B. B
- C. C
- D. D

Sample Items from the 10th/11th Grade U.S. History MCAS Test

- 31** In 1937, after four years of steady growth, economic activity in the United States declined by 27 percent. Which of the following terms **best** describes this decline in the economy?
- A. affluence
 - B. inflation
 - C. recession
 - D. recovery

- 32** The poster shown below is from World War II.



J. Howard Miller, "We Can Do It!" (ca. 1942)
© J. Howard Miller/CORBIS

This World War II poster was made to encourage women to

- A. buy war bonds.
- B. enlist in the military.
- C. plant victory gardens.
- D. work in defense industries.

Sample Items from the 10th/11th Grade U.S. History MCAS Test

- 33** Which of the following issues was central to the Nullification Crisis of 1832-1833?
- A. due process
 - B. laissez faire
 - C. states' rights
 - D. women's rights
- 34** Which of the following is a responsibility of the Massachusetts Supreme Judicial Court?
- A. to enforce laws
 - B. to write legislation
 - C. to hear cases on appeal
 - D. to propose new amendments
- 35** Which of the following is the **main** reason President Harry Truman gave for ordering an atomic bomb to be dropped on Hiroshima?
- A. to show the world the horror of atomic bombs
 - B. to avoid large American casualties in an invasion of Japan
 - C. to persuade communist China that it should stay out of the war
 - D. to demonstrate to Germany that the United States had an atomic bomb

- 36** The excerpt below is from George Washington's Farewell Address in 1796.

It [party conflict] serves always to distract the public councils and enfeeble the public administration. It agitates the community with ill-founded jealousies and false alarms; kindles the animosity [hatred] of one part against another; foment[s] [provokes] occasionally riot and insurrection.

— George Washington,
Farewell Address (1796)

Which of the following statements **best** summarizes George Washington's view of political parties?

- A. They were sources of corruption.
- B. They were controlled by social elites.
- C. They were sources of division in the country.
- D. They were prone to influence by foreign powers.

NOTES

ⁱ Studies that correlated teachers' orientations with students' historical literacy were sought out but not found. Thornton's (1991) review of the literature corroborates this absence of empirical work and notes that "Although they endorse a chronological approach, neither the NCSS (National Commission on Social Studies) nor Ravitch and Finn present any research evidence to support their assertion. As Levstik (1990b) points out, the limited research available suggests that approaches other than chronological (e.g., more topical or thematic approaches) may be more effective in engaging students" (Thornton, p. 236).

ⁱⁱ Most education philosophers take "primary education," to subsume the years of kindergarten through high school.

ⁱⁱⁱ Indeed, Kelly et al (2007) argues many historians and the field as a whole, "traffics" in all of these categories.

^{iv} One doctoral dissertation employed survey methodology to explore prevalence of traditions based upon Barr, Barth and Shermis' "Social Studies Preference Scale." See Andres (1982) also see Landman (2000).

^v For topics on multicultural citizenship see Banks (2004), Gay (1997; 1996), Grelle and Metzger (1996), Kaltsounis (1997), Parker (1997). For topics related to discussing controversial subjects in HSS classrooms see Hess (2005; 2008). For topics related to social transformation see Stanley (2005). For topics related to trends of inclusion and perspective taking see Thornton (1991). For topics related to integration of traditions see Cornbleth (1985). For topics related to character education see Hoge (2002). For topics related to critical thinking see Metzger and Marker (1992). For topics related to issues centered curriculum see Stanley & Nelson (1994).

^{vi} Another oft-cited source is Popham (2001) "The truth about testing: An educator's call to action."

^{vii} In fact, in the Ravitch and Finn (1987) book, the authors state clearly, "We do not assert that American youth know less about the past than their predecessors. This may be true, but one cannot verify it from the data presented in this book. This assessment was administered once, and there are no previous test results with that it can be compared. Based on the findings reported here, we are not able to state that history and literature in the schools are taught (and learned) either better or worse than they used to be. We simply don't know" (p. 200).

^{viii} It should be noted that the history standards were revised and rereleased in 1996 in part due to some of the criticisms raised included in Stern's (1994) piece.

^{ix} To clarify, this was an empirical study of the departmental effect of the MCAS history test (dissertation).

^x This will be done once the dissertation is finalized.

^{xi} A firsthand account of the experience of being on an Assessment Development Committee was given by a department chair of one of the participating schools in this study.