

Fall 1998

Differences in achievement across item types on the advanced placement United States history examination: The relationship of sex, gender identity, cognitive orientation, and conception of the nature of historical knowledge to domain-specific academic performance

Fred J. Wetzel

University of New Hampshire, Durham

Follow this and additional works at: <https://scholars.unh.edu/dissertation>

Recommended Citation

Wetzel, Fred J., "Differences in achievement across item types on the advanced placement United States history examination: The relationship of sex, gender identity, cognitive orientation, and conception of the nature of historical knowledge to domain-specific academic performance" (1998). *Doctoral Dissertations*. 2051.

<https://scholars.unh.edu/dissertation/2051>

This Dissertation is brought to you for free and open access by the Student Scholarship at University of New Hampshire Scholars' Repository. It has been accepted for inclusion in Doctoral Dissertations by an authorized administrator of University of New Hampshire Scholars' Repository. For more information, please contact nicole.hentz@unh.edu.

INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA
313/761-4700 800/521-0600

DIFFERENCES IN ACHIEVEMENT ACROSS ITEM TYPES ON THE ADVANCED
PLACEMENT US HISTORY EXAMINATION: THE RELATIONSHIP OF SEX,
GENDER IDENTITY, COGNITIVE ORIENTATION, AND CONCEPTION OF THE
NATURE OF HISTORICAL KNOWLEDGE TO DOMAIN-SPECIFIC ACADEMIC
PERFORMANCE

BY

Fred J. Wetzel
BA, Colby College, 1965
MBA, State University of New York at Buffalo, 1971

DISSERTATION

Submitted to the University of New Hampshire
in Partial Fulfillment of
the Requirements for the Degree of

Doctor of Philosophy

in

Education

September, 1998

UMI Number: 9907593

**UMI Microform 9907593
Copyright 1998, by UMI Company. All rights reserved.**

**This microform edition is protected against unauthorized
copying under Title 17, United States Code.**

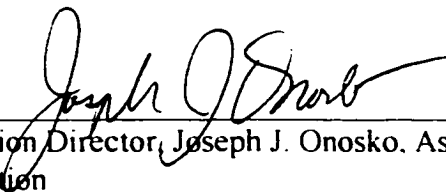
UMI
300 North Zeeb Road
Ann Arbor, MI 48103

ALL RIGHTS RESERVED

c 1998

Fred J. Wetzel

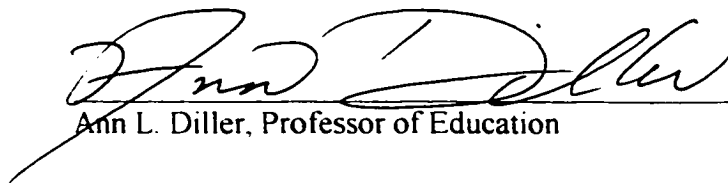
This dissertation has been examined and approved.



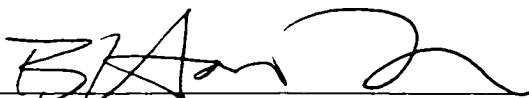
Dissertation Director, Joseph J. Onosko, Associate Professor
of Education



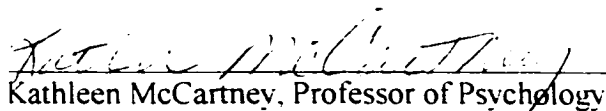
Richard M. Barton, Assistant Professor of Education



Ann L. Diller, Professor of Education



Barbara E. Houston, Professor of Education



Kathleen McCartney, Professor of Psychology

June 8, 1998

Date

DEDICATION

To Mom

When it comes to knowledge ... she knows

ACKNOWLEDGEMENTS

Every story has a beginning, and Bev Miyares was there at the start to ask the question for the first time. Others helped along the way with their own stories ... Karen Laba was there for me and others all along the way. Organizations can contribute their stories too. The College Board helped with tuition (and sometimes fees ...), time and trees. AP friends were always happy (or at least seemed to be) to hear the latest reports on writes, rewrites, and more rewrites.

Without the committee I had the story would have been so much less (when does a story cease to be itself and become another ... ?). Kathleen McCartney has intelligence and grace and modeled both beautifully. Ann Diller can support and ask for more at the same time (maybe that's what support is ... ?) and seems to be enjoying it the whole while. Barbara Houston daunts, but for the best of reasons. Rick Barton and Joe Onosko went far far beyond the call. They supported someone who should have done this years ago without fuss and always with encouragement. Funny how you want to do well when you feel others want you to do well too. The beer will come soon.

The story started when Lindsey went to college six falls ago. I thought a final try at this doctorate business would help ease the transition (no, I won't say loss ...). Little did I know how full the time would be. And through it, she, now a real live teacher-person herself, was always a supporter - maybe wondering just why I was doing it, but maybe just a little proud as well. Hope looked across the table too many times and smiled while I was writing. I noticed - I really did (I was just pretending to be hard at work revising chapter 3).

And now as the final 1/8 inch on the margin of the figures is adjusted for binding, I say thanks to all. A story of a lifetime - for me at least, and that's what it's all about, isn't it?

TABLE OF CONTENTS

DEDICATION	iv
ACKNOWLEDGEMENTS	v
LIST OF TABLES	x
LIST OF FIGURES	xi
ABSTRACT	xii

CHAPTER		PAGE
I.	CHAPTER 1	1
II	CHAPTER 2	13
	<u>Development in Infants and Children</u>	14
	<u>The psychoanalytic perspective.</u>	15
	<u>Social learning.</u>	18
	<u>Cognitive-developmental approaches.</u>	22
	<u>Adolescent and Post-Adolescent Cognitive Development</u>	26
	<u>Knowledge as received certainties.</u>	29
	<u>Knowledge as relative beliefs.</u>	31
	<u>Knowledge as conditional constructions.</u>	34
	<u>The Work of Gilligan and Bruner</u>	38
	<u>Gilligan and moral thinking.</u>	39

	<u>Bruner and modes of cognition</u>	40
	<u>Achievement and Ability</u>	45
	<u>Multiple-Choice Testing and Gender</u>	53
	<u>The Assessment of Achievement in History</u>	62
	<u>Summary and Conclusion</u>	66
III.	CHAPTER 3	71
	<u>Sample</u>	71
	<u>Instrumentation</u>	74
	<u>Nature of Historical Explanation and Truth questionnaire</u>	74
	<u>Narrative/Paradigmatic Organization of Knowledge</u>	
	<u>questionnaire</u>	79
	<u>Bem Sex-Role Inventory</u>	77
	<u>Student Biographical Questionnaire</u>	79
	<u>The Advanced Placement examination in US History</u>	80
	<u>Procedure</u>	82
	<u>Data Analysis</u>	85
	<u>Descriptive statistics</u>	86
	<u>Intercorrelation of data</u>	87
	<u>Multiple-regression models</u>	88
	<u>R²-change models</u>	89

IV.	CHAPTER 4	91
	<u>Descriptive Statistics</u>	91
	<u>Performance on the AP US History examination.</u>	91
	<u>Sex and Bem category.</u>	91
	<u>Academic achievement.</u>	94
	<u>Disposition toward History.</u>	95
	<u>Nature of historical truth.</u>	96
	<u>Cognitive orientation.</u>	97
	<u>Intercorrelation of Data</u>	97
	<u>Predictor variables.</u>	97
	<u>Predictor and outcome variables.</u>	98
	<u>Multiple Regression Models</u>	102
	<u>Total score.</u>	103
	<u>Multiple-choice.</u>	103
	<u>Document-based question.</u>	105
	<u>Standard essays.</u>	105
	<u>R²-Square Change Models</u>	106
	<u>Exogenous factors and intermediate variables.</u>	106
	<u>Total score.</u>	112
	<u>Multiple-choice.</u>	112
	<u>Document-based question.</u>	113
	<u>Standard essays.</u>	113

V.	CHAPTER 5	114
	<u>Overview of Findings</u>	114
	<u>Descriptive characteristics of students in the sample.</u>	115
	<u>Cognitive orientation.</u>	116
	<u>Nature of historical truth.</u>	119
	<u>School and PSAT scores.</u>	124
	<u>R² in multiple-regression and R²-change models.</u>	128
	<u>Overview summary.</u>	129
	<u>Limitations of the Study</u>	130
	<u>Recommendations for Further Research</u>	132
	<u>Cognitive orientation.</u>	132
	<u>Epistemological development.</u>	133
	<u>School and school climate.</u>	135
	<u>Conclusion</u>	137
	REFERENCES	139
	APPENDIX A	147
	APPENDIX B	152
	APPENDIX C	155
	APPENDIX D	165
	APPENDIX E	167
	APPENDIX F	168

LIST OF TABLES

Table 1	Sex and Bem-Category of Subjects
Table 2	Comparison of Female and Male Distributions Across Categorical-Level Descriptive Variables
Table 3	Comparison of Female and Male Distribution Across Interval-Level Descriptive, Achievement, Disposition, and Outcome Variables
Table 4	Correlations Among Predictor Variables (All Subjects)
Table 5	Correlations Among Predictor Variables (Males and Females)
Table 6	Correlations Among Predictor and Outcome Variables
Table 7	Multiple-Regression Partial Correlation and Beta Coefficient Values for Predictor Variables Across Four Outcome Variables
Table 8	R^2 and R^2 -Change Values for Four Multiple-Regression Models Across Four Outcome Variables

LIST OF FIGURES

- Figure 1** Model representing influence of exogenous and intermediate variables on AP US History achievement. Outcome variable: Total Score
- Figure 2** Model representing influence of exogenous and intermediate variables on AP US History achievement. Outcome variable: M-C Score
- Figure 3** Model representing influence of exogenous and intermediate variables on AP US History achievement. Outcome variable: DBQ Score
- Figure 4** Model representing influence of exogenous and intermediate variables on AP US History achievement. Outcome variable: Essay Score

ABSTRACT

DIFFERENCES IN ACHIEVEMENT ACROSS ITEM TYPES ON THE ADVANCED PLACEMENT US HISTORY EXAMINATION: THE RELATIONSHIP OF SEX, GENDER IDENTITY, COGNITIVE ORIENTATION, AND CONCEPTION OF THE NATURE OF HISTORICAL KNOWLEDGE TO DOMAIN-SPECIFIC ACADEMIC PERFORMANCE

by

Fred J. Wetzel

University of New Hampshire, September, 1998

This study explored the reasons why males outperform females on the AP US History examination, a three-part, college-level test taken by large numbers of high-achieving grade 11 students each year. Males tend to outperform females on the overall test, and on the multiple-choice section. Females show relatively stronger performance on the examination's document-based question (a section that requires analysis of original historical documents). Performance on the standard-essay part of the test does not consistently favor either sex.

Moving beyond explanations that rely on sex-related generalized differences in ability and interest, the study explored the possibility that performance was dependent on conceptions about the nature of historical truth and characteristic ways of organizing new information. It was hypothesized that males and females think about historical truth and organize information differently, and that these differences would provide relative advantages on the three different types of items used on the AP US History examination.

Cluster sampling of intact classes from public high schools in two New England

states (MA and NH) resulted in a sample of 190 students (106 females and 84 males) from 14 schools. Students selected for the sample completed four instruments: (a) the Nature of Historical Explanation and Truth questionnaire; (b) the Narrative/Paradigmatic Organization of Knowledge questionnaire; (c) the Bem Sex-Role Inventory; and (d) the Student Biographical questionnaire. The overall score and scores for the three sections of the 1997 AP US History examination served as outcome variables in the study.

Nineteen hypotheses which focused on possible sex and gender-linked conceptions of historical truth and cognitive orientation and differences in performance by sex on the AP US History examination were tested. Analysis of the data showed both female and male students to be academically able and positively disposed toward history. Though students in the sample showed evidence of advanced epistemological development, neither cognitive orientation nor epistemic disposition (conceptions of historical truth) was found to be a strong predictor of performance on the AP US History examination. For both males and females, examination performance was most closely related to PSAT scores and the school attended.

**THE RELATIONSHIP AMONG STUDENT SEX, GENDER IDENTITY, AND
CONCEPTIONS OF KNOWLEDGE TO ACHIEVEMENT DIFFERENCES IN
HISTORY**

Chapter 1

Male and female high school students who take Advanced Placement courses have similar academic profiles: relative to their peers, they are academically able, high achieving, and well motivated. Yet despite this similarity, a difference in performance across sex on AP examinations has existed for many years. Males have traditionally done better than females on AP examinations in Mathematics, while females achieve higher scores on tests of languages such as French and Spanish. Though the differences may decrease as more and better-prepared students of both sexes enroll in AP courses, the current performance differences do little to refute the conventional wisdom that boys are more mathematically oriented, while girls possess superior verbal skills.

Scores on one Advanced Placement examination show particularly interesting sex-related performance differences. For the AP US History examination, with a nearly equal distribution of male and female test-takers (in recent years females have slightly outnumbered males) males have consistently received higher scores than females. This is perplexing given the nature of history as a discipline and the AP course in US History in particular. Becoming proficient in history requires much reading and writing, activities that are conventionally seen as favoring females. Yet males clearly outperform females on the culminating assessment for the AP US History course. Why such a difference in performance occur? Does it represent real differences in ability and achievement, or could it be due to other contextual factors?

Before attempting to answer this question, it is important to be clear about what I mean by difference in this particular context. Let me begin the process of conceptual clarification by telling a story. Five years ago, during a visit with the Commissioner of the Massachusetts Department of Education, I was asked if I knew why males in the state did so much better than females on the AP US History examination. I not know why -- I had not even known there was a difference in performance. My conceptual lens had focused on differences in the scores of females and males in other AP subject areas such as mathematics and world languages, but because I was not expecting to find differences in achievement in history, I simply had not paid attention to the scores on the AP US History examination. Yet not only was there a difference favoring males, but it was the largest of any difference favoring either females or males across all the AP examinations given in Massachusetts that year.

Now curious, I called the Educational Testing Service (the organization that administers AP examinations for the College Board) and asked for more psychometric information about scores on the AP examination in US History. The examination includes items of three types: multiple-choice questions, essay questions, and a document-based question (which requires students to consider information contained in a set of historical documents before responding to a relatively open-ended essay question). The statistical information from ETS showed that males as a group outperformed females by a wide margin on the multiple-choice questions, that there was essentially no difference in the scores on the standard essay questions, and that females as a group did slightly better than males on the document-based essay question. Because males outperformed females so

much more on the multiple-choice questions (which were also weighted more heavily in the determination of the final score) than females outperformed males on the document-based question, males received higher overall scores on the test than females.

In a simple way the question had been answered: males received higher scores than females on the AP US History examination because they performed much better on questions given in the more heavily weighted multiple-choice format. However, rather than solving the problem, this simple answer led to the more complex and compelling question: why would student performance on different types of test items be associated so closely with their sex, especially for students who are so similar academically?

One possibility could be that performance on different item types is congruent with developed verbal and, perhaps to some extent, quantitative ability. For example, it might be that students who have more highly developed verbal skills are better document-based essay writers; students with high mathematical reasoning ability might be found to perform well in answering more closed-ended, multiple-choice test items. If these developed abilities were found to exist disproportionately in one sex or the other, a more interpretively elaborate answer to the question could be developed by showing the connections among these reasoning abilities, the sex of the student, and performance on different types of test items.

Another possibility could be that differences in achievement are due to selection variables. For example, high achieving girls might choose not to enroll in the AP US History course, or, once enrolled in the course, to not take the AP examination at the same rate as males.

A third possibility for explaining differences in achievement--and one I actually explored soon after becoming aware of the difference in performance--was that student performance was reflective of some type of gender(ed) interaction with their teachers. Same-sex pairings of teachers and students (e.g., female teachers and female students) might be associated with learning and subsequent test performance that was different from the learning and performance of students who were matched in a cross-sex pairing with their teacher.

I subsequently examined (Wetzel & Barton, 1995) the possibility of a relationship between test performance and membership in one of the four possible teacher/student, sex-paired categories (e.g., female teacher/male student). There was no evidence of the relationships (and statistical interactions) I thought might exist, though the data did show that all students (male and female) who were taught by female teachers did better on the examination as a whole than students taught by male teachers. The finding was provocative (are women better history teachers than men ... ?) but seemed to suggest more about the importance of quality teaching than about the performance on different item types for students of different sexes.

For the time being, I put the matter aside. The following semester, after reading an assignment about the process of making moral decisions (Noddings, 1984), I began thinking that perhaps the investigation into differences between women's and men's moral thinking could help to explain sex-related differences in academic performance. Carol Gilligan's work was particularly important in this regard (Gilligan, 1982/1993; Gilligan, Ward, & Taylor, 1988). Her ideas about differences in cognitive development between

females and males, which was the grounding for the ideas expressed and carried forward by Belenky and her colleagues in Women's Ways of Knowing. (Belenky, Clinchy, Goldberger, & Tarule)¹ and her claims for differences in approaches to moral decision-making by sex (with females more closely associated with an ethic of care and males an ethic of justice) led me to consider the possibility that the process of developing criteria for thinking about historical truth (what I came to call truth-in-knowledge) and the development of cognitive orientation might be mediated by sex and gender.

This story about differences in performance on the AP US History examination does not finish with the comparison of female and male performance on each section of the AP US History examination. Rather, as with any intriguing story, it leads to other questions -- in this case, questions about what is contributing to the observed differences in academic performance. The search for answers to these questions begins with this hypothesis: sex-related performance on the three types of test items that make up the AP US History examination is associated with how students think about the criteria for determining historical truth (truth-in-knowledge), and how they actually organize information into knowledge claims (cognitive orientation).

With regard to the part of the hypothesis that involves criteria for thinking about historical truth, I am interested in what students think about the groundings of knowledge -- specifically, what they think determines the truth of a knowledge claim or, even more specifically, if a knowledge claim can entail the idea of truth. This is similar to what Baxter-Magolda (1992) called beliefs about the nature of knowledge and to what King

¹ Truth-in-knowledge was the first construct I assumed to influence performance on the AP US History examination. The second was cognitive orientation, a concept developed by Bruner (1985).

and Kitchner (1994) considered a person's concept of knowledge and the attendant process of justification. Furthermore, in keeping with a basic tenet of feminist approaches to epistemology, I will consider these ideas as they are found in a specific context. In this case, the contextual setting is the learning of US History at the AP level in high school.

It would be an open (and highly debated) question if I were to speculate whether individuals move along a developmental arc in thinking about historical truth. I do not intend to enter the debate over developmental stages in this study. Though I feel there are developmental processes at work, it is sufficient for the purposes of this study to posit that individuals think about historical truth in three different ways. The most I would claim about the process of creating a concept of historical truth is that it is developmental in a loose sense. That is, it seems reasonable that when individuals begin to learn about something (particularly young or otherwise inexperienced learners) they are more likely to consider the knowledge in question to be external to them and to exist in a fully-formed state. I also believe when individuals think about the origin of historical truth, they produce descriptions that are sufficiently different from one another so that the descriptions are clearly recognizable as falling into distinct categories. Again, the categories do not necessarily represent developmental stages, but they are categories nevertheless, and therefore distinctive in their identifying characteristics.

I want to point out at this time that while these categories for thinking about the nature of historical truth may appear very similar to the epistemological categories that have emerged in the wake of the ways-of-knowing literature, what I am thinking about is probably better characterized as a group of sub-categories within a broader ways-of-

knowing schema. For example, while a ways-of-knowing approach to making meaning might discuss the importance of power relationships in determining who has the right to know, I do not focus on such inputs and the outcomes of such relationships (e.g., who may ultimately have access to knowledge and who will not). Rather, I am concerned with a partial outcome of the overall process of which power relationships are a determining factor -- an outcome which can be described as what an individual characterizes as the source and even the possibility of claims for truth. I believe these conceptions about the source of knowledge claims can be arrayed descriptively (if not developmentally) as objective and certain and external to the knower, through subjective and arising from within the knower, to a position in which truth is considered to be socially constructed and broadly contextual.

Moving from theory to practice, and using the concept of historical truth as an example, it could be that a person who believed historical truth was absolute would find it easier to demonstrate knowledge when the assessment was designed to elicit one clearly preferred (correct) response. Such would be the case with items presented in a multiple-choice format. A person who believed that historical truth is a less than permanent outcome of human intellectual activity, could find it harder to work in a multiple-choice question format, but more comfortable when asked to create a response containing elements of narrative. It is not unreasonable to think such a person would be able to demonstrate what they knew better in the second assessment situation than in the first.

It may also be the case (as I will try to show in the next chapter in the discussion of sex and gender development) that beliefs about concepts such as historical truth are

themselves related to sex and gender, with the result that conceptions of historical truth could be related to assessment performance differently for different sexes and genders.

As for the part of my hypothesis that refers to how information is organized, this process is almost pre-epistemological in the sense that it supposes a naturally preferred way of structuring and ordering perceptions before the more conventional epistemological activities (e.g., justification) are considered by the learner. Interestingly, and as will be shown when I say more about Bruner, there is no claim that this process of organizing information is developmental. There is no hint in Bruner's writing that preferences for organizing information develop first in one way and then change form. Bruner is silent on the manner of how the preferences develop and he offers no indication that a preferred manner of organizing perceptions will change over time.

The idea that there are characteristic ways of organizing information makes a connection with epistemology when it is considered in relation to Kornblith's (1994) argument that people act as if they are capable of constructing or acquiring knowledge and are able to make use of it once created or gained. My move would be to go beyond Kornblith's claim and state that as part of the process of constructing or acquiring knowledge, we use naturally occurring and preferred ways of organizing the information available in our environments.

Turning directly to Bruner (1985), the idea that individuals show characteristic ways of cognitive organization has led him to suggest that we prefer one of two modes of thought, which he has identified as narrative and paradigmatic. These two ways of thinking direct our cognitive focus and guide the active and selective perception of our

natural and social environment. Bruner also claimed these two cognitive modes operate independently and across all disciplines of inquiry.

Paradigmatic thinking appears to operate along the lines of concept formation. The operations of paradigmatic thinking are those by which categories are established and connected to form systems of meaning. Paradigmatic thinking seeks universals and moves quickly to abstract representation. Narrative thought, on the other hand, deals with human (or human-like) intention in the particular and the resulting action and outcomes of such intention.

The outcomes of paradigmatic thought can be logically falsified. In fact, applying rules of conventional logic is the way that paradigmatic thinking is validated. Narrative thinking is characterized by under-determined verisimilitude. That is to say, the criterion for high quality narrative thinking is the extent to which the outcome has a sense of reality that resonates with the ideas and feelings of those listening to or otherwise receiving the message. A good narrative thought may or may not possess conventional veridical reality.

Bruner does not speak to differences in cognitive organizing across sex or the constructed categories of gender, but has much to say about characteristic ways of making sense of the world. His claim--that two cognitive modes account for how we perceive the natural and social world and how we organize the content of what we think about--provides a powerful heuristic for linking thinking and learning and subsequent performance on measures of assessment such as examinations.

Again moving from theory to practice, using the idea of cognitive orientation, we might say that a person who employed a narrative mode of organizing information would

be expected to show a greater ease in demonstrating knowledge when the assessment itself was organized narratively (such as is the case with the document-based question on the AP US History examination). A person using the paradigmatic mode might perform better when the assessment asked for the identification of the one best answer, separate and distinct from other possibilities, an answer that was in some way attainable through a process involving logical exclusion of competing possible answers (such a process begins to look like what could be involved in answering a question posed in a multiple-choice format).

Having offered a limited explication of what could be thought of as the two major independent variables in my research hypothesis, I want to argue again that both need to be considered when attempting to explain differences in academic performance across sex and gender. Taken together, (1) beliefs about the nature of historical truth, and (2) preferences regarding the organization of information, will provide the explanatory framework to guide research into the differences in performance between females and males on the AP examination in US History.

At this point it is necessary to raise two additional issues which may help in the search for understanding sex-related differences in performance on the AP US History examination. The first issue is the need to distinguish between sex as biologically-based and gender as constructed from the interaction of biological and cultural factors. It may not be simply the biological sex of the student that is related to differences in academic performance, but rather a more complex, relationally-constructed sense of gender. It may be that factors related to self-identified conceptions of gender have a greater influence on

achievement than distinctions based solely on traditional biological categories of male and female. Therefore, the study needs to consider performance outcomes using a measure of gender for those in the study sample as well as looking at results using the more conventional variable of biological sex.

The second issue is that the structure and content of what is being studied (in this case, the discipline of history -- specifically, US history) may itself play a part in learning and subsequent achievement. As individuals, we may or may not organize information the same way across disciplines (Bruner gives no indication that a person's organizing orientation would be different in different contexts), but since the research at hand involves a specific discipline, I feel it is better to acknowledge that grounding and focus the inquiry on ideas concerning historical knowledge and historical truth. Since a more complete explication of the concept of cognitive orientation might show it to differ does across different disciplines, research into performance differences is better to the extent that it recognizes the discipline in which performance is being assessed. To use history as an example, definitional conventions such as what counts as historical fact, structural elements such as the nature of historical questions, and matters of process such as the relationship among elements of historical inquiry, all could influence learning and achievement in history. Different discipline-specific conventions, elements, and processes could control the discourse in other academic areas.

There has, in fact, been curiously little research conducted on sex and gender-related achievement differences outside mathematics and the physical/biological sciences. This may be due to the perceived economic importance of achievement in mathematics

and science. It is also possible that a conventional wisdom underlying the process of problem-posing is at work; that is, there may be a sense that girls and women are equally if not more capable than males in linguistically-based domains such as the humanities. Yet this characterization, too, may overstate differences in achievement between the sexes in these domains. After all, while girls outperform boys on AP-level foreign language examinations, boys have done markedly better than girls in AP US History for many years.

Thus, within one academic domain, the consideration of two factors -- beliefs about the nature of historical truth and characteristic ways of organizing information -- generates a richer conceptualization of the research problem, making it one of determining the nature and extent of the relationship between epistemological and cognitive beliefs and discipline-specific achievement measured by different types of assessment items. The specific mechanisms involved in the effects of each of these factors and their possible interaction undoubtedly will remain to be determined. For this study I will treat the factors individually, and hypothesize that for Advanced Placement US History, differences in student conceptions about the nature of historical truth and differences in modes of organizing information will show, for different types of test items, distinctive patterns of academic performance across sex and self-identified conceptions of gender.

The ultimate goal of the research is to contribute to understanding the complex interplay among the student, the curriculum, and other elements of the learning environment. With an increase in knowledge about the effects of individual elements and their relationships, the design and delivery of the curriculum can be better shaped so that interest and achievement in history will be improved for all students.

Chapter 2

The purpose of this chapter is twofold. As a literature review, the intent is to document the work of researchers who have considered sex and gender-related issues in education. In addition, the claims and arguments in the chapter will serve to construct a major portion of the study's conceptual framework. By showing how researchers have thought about the development of sex and gender identity and have considered how sex and gender identity may influence learning and thinking about knowledge, I intend to show the need for research to examine the following propositions: (a) that students conceptualize the nature of knowledge (specifically the nature of truth-in-knowledge) and organize information in ways that are associated with gendered development, and (b) that these differences in epistemic and general cognitive orientation are associated, in turn, with differences in academic achievement.

The literature review begins with a consideration of how gender roles are developed in infancy, with particular attention given to findings which suggest possible connections between early development and later gender-contingent differences in conceptions of the nature of truth-in-knowledge claims. Two models of cognitive functioning, drawn from the work of Gilligan and of Bruner, are then examined for their contributions to understanding epistemic and general cognitive orientation. Gilligan's work on differences in reasoning between women and men, though challenged by some, is important in providing the possibility that gendering processes in the culture can

differentially affect how males and females both think about claims for truth and cognitively organize information.

The chapter concludes with a three-part consideration of achievement and testing. The first section is an overview of the relationship between sex and academic achievement. The second part considers the more specific issue of sex-linked differences in achievement when testing is carried out in a multiple-choice format. The third part examines the relationship between sex and achievement in the domain of history.

Overall, I am proposing a conceptual model to account for differences in achievement between females and males in one academic domain. The model claims that a student's characteristic way of organizing information (i.e., narrative or paradigmatic cognitive orientation) directly affects academic performance depending on the type of assessment. It also claims that academic performance is influenced by the manner in which a student considers the nature of historical truth. Finally, the model also claims that there is an indirect effect of cognitive orientation on achievement; that is, cognitive orientation is mediated through beliefs about the nature of truth-in-knowledge (whether truth is seen as absolute, relative, or constructed) to account for characteristically male and female patterns of thinking and achievement. The subsequent analysis of data in the study will show the effects of both these direct and indirect influences on achievement.

Development in Infants and Children

Approaches to development can be divided into several theoretical groupings, and regardless of the relative importance researchers ascribe to essentialist and developmental processes, they seem to support the proposition that, from infancy onward, development

taken as a whole is a process that has heavily gendered overtones. For example, Hudson and Jacot (1991), in their account of male development, looked at the collective effects of biological, psychological, and social forces on the formation of gender identity in males. Beal (1994), in her work on development, focused on distinct sets of psycho-cultural factors. She therefore chose to examine the development of gender identity from the psychological and cultural perspectives of psychoanalysis, social learning theory, and cognitive-developmental theory. Though it can be argued that gender development is primarily driven by physical differences between the sexes, I find Beal's argument for a more dynamic and less biologically deterministic approach to development more convincing and therefore will use the three psycho-cultural categories proposed by Beal (1994) to group ideas about how gender identity develops in pre-adolescent children.

The psychoanalytic perspective. In the traditional Freudian approach to gender development, there is assumed to be little differentiation between male and female infants through the oral and anal stages. Parents, especially mothers, are seen as instrumental for the emergence of Ego in the pre-Oedipal stage, but the effort to assist in the channeling of the child's psychic energy (the Id) into the development of a socially-accepted and socially-constructed problem-solving Ego (and later, Superego) is not directed more to one sex than the other. Only when the child reaches the genital stage at ages three to six do developmental events play a major part in defining gender. For boys, whom Freud saw as having a more direct route to development (though the conditions contributing to development for boys might be more traumatic than those for girls), the Oedipal crisis, precipitated by a sexual desire for the mother with the awareness of the foreshadowed

danger this desire will entail, means moving away from the mother as the most influential role model and identifying, in a self-protecting way, with the father. The path is somewhat less likely to lead to definite resolution for girls, with their first turning to the father for support and only later returning to the mother in re-identification and adopting the feminine role. The renewed identification with the mother is always somewhat ambivalent, with the result that the Superego doesn't become as fully developed for girls.

Although much has been written about the weakness of psychoanalysis as a basis from which to develop a general approach to gender differentiation (Beal, 1994), a psychoanalytic perspective nevertheless provides one grounding from which to develop an understanding of how females and males may differ in their general cognitive activity. Freudian concepts such as developmental stages have provided points of departure for theorists such as Piaget and other researchers with interests in cognitive developmental issues. Freud also gave credence to the idea that early experiences could have long-lasting influences. In addition, although it is often overlooked, the conception of object identification with like and unlike-sex parents emphasizes the role of learning in the development of gender.

Psychoanalytic theory makes several claims about differences in gendered development which could contribute to the development of different epistemic stances for males and females. Chodorow's re-interpretation of psychoanalytic theory is important in this regard. Chodorow (1989) examined the overall nurturing and caretaking role of mothers, and concluded that a mother views a boy-child, even in the pre-oedipal stage, with a sense of what is different, while identifying with (almost as an extension of herself)

what is similar in a girl-child. While it is important to remember the influence of broad cultural influences on psychological development (consider both the sexist proscription in our culture against young males being known as “mama’s boys”, and Chodorow’s statement that these engendering attitudinal differences are occurring in “Western middle-class women”), her writing nevertheless has a strong tone of universality to it. For example, Chodorow wrote, “... a mother unconsciously and often consciously experiences her son as more of an ‘other’ than her daughter. By contrast, the female’s self is less separate and involves ... difficulties with a sense of separateness and autonomy. Girls grow up with a sense of continuity and similarity to their mother, a relational connection to the world” (Chodorow, 1974, p. 48). Thus, using the psychoanalytic explanation that resulting ego boundaries are less well developed for girls than boys, Chodorow set the stage for girls, more so than boys, coming to define themselves as in relation to others.

Arguing from the perspective of gender development in males, Hudson and Jacot (1991) claimed that the additional step of counter-identification with the father is more difficult for boys than the return to the mother as object choice for girls. Dis-identification with the mother and subsequent counter-identification for boys, which can take place at a relatively young age, result in what Hudson and Jacot call the male “wound”, and they take the approach that much of what follows in the way of development for males follows directly from it. Girls do not experience the developmental consequences of separation as much as boys. As a result, males are more likely to draw clear distinctions between the personal and the impersonal. Males are also more likely to develop a preoccupation with

issues of intellectual control. In terms of style, males are likely to be intolerant of arguments that appear intuitive and indeterminate, and lack formal structure (Hudson & Jacot, 1991). Male style emphasizes the virtues of dispassion and objectivity, applied in a world of things, mechanisms, abstract ideas, and systems. Males will more often cast arguments in terms of dualities and dialectical oppositions, and prefer ideas which have a clear nature of truth (Hudson & Jacot, 1991).

Although proponents of psychoanalytic approaches to development should not be thought of as essentialist--they do not assume that the process of personality development is dictated solely by genetic mechanisms (remember the social context caution of Chodorow and the socially-provided idea of a father in counter-identification) -- one is nevertheless left with a sense that under a Freudian conception of gender development the factors and processes affecting behavior are less open to modification from forces beyond or outside the immediate familial sphere of the individual child.

Social learning. The belief that influences both within and outside the bond of individual and family can shape cognitive activity and personality is developed in theories which assign particular importance to the role of social learning. The key tenet in this approach is the idea that infant and child behavior will elicit different reactions from parents and other adults according to culturally determined ideas of appropriate behavior for a child. The biological sex of the child is often used as a natural distinguishing characteristic to guide such selective responses to behavior (Beal, 1994). Regarding the development of gender identity, social learning theories incorporate beliefs about learning

found in the associational and behavioral theories that dominated learning theory until the development in the 1970's of the more purposeful cognitive approaches toward learning.

Using the behavioristic concept of reinforcement and the social-psychological idea of role models, social-learning theory holds that the differential reinforcement of behavior by family members and the broader society shapes the emerging concept of self and gender in infants and young children. In a given culture, conventional gender-type behavior will be recognized and rewarded -- cross-gender behavior will not. What counts as appropriate gender-type behavior is culturally determined in the same manner as other culturally-defined constructs. It is therefore as mutable and evolutionary in character as other cultural constructs.

The occurrence of social learning is most visible in specific areas of behavior such as play, independence, and achievement orientation (Block, 1983; Fagot & Hagan, 1991; Huston, 1983). Much of this activity has gendering overtones which has implications for subsequent behavior. For example, researchers have found that in our culture, boys are granted more time to be on their own, while girls are kept under stricter familial supervision (Beal, 1994). Though this does not necessarily mean that girls are therefore more likely to participate in group-centered and interactive social activity than boys, it would seem reasonable to consider the possibility of such an outcome.

In the area of achievement, research has shown that parents, when reading to and questioning young children, are less likely to encourage girls to arrive at a solution to the questions posed and more likely to emphasize the social situation of being together and enjoying the event (Weitzman, Birns, & Friend, 1985). Fathers show less warmth and use

more direct and intrusive teaching strategies when working with boys than girls. Though very young infants show a similar repertoire of behaviors, male infants are rewarded more for efforts to use physical assertion, while female infants are reinforced for their attempts to communicate verbally (Fagot, Hagan, Leinbach, & Kronsberg; 1985).

In the research cited above, it is clear that young female children are being treated differently than young male children. I will not compare social learning with traditional psychoanalytic claims (psychoanalysis holds to a lack of gendering activity before the genital stage). And though it is not necessary for my research that possible discrepancies between the approaches be reconciled, the difference between the two approaches to development should be noted.

In general, the studies described in Beal (1994) show a pattern of expectations for boys to become independent while girls are allowed (and even encouraged) to remain, if not dependent, at least much closer to their parents. The strong avoidance of boys to traditional feminine activity (stronger than the corresponding avoidance by girls to male activity) might well be determined by the strength of the messages to boys, and the presumed behavioral incorporation of the messages, regarding appropriate gendered behavior (Bussey & Perry, 1982).

In educational settings, the research describes a social dynamic of differential shaping of behavior and beliefs based on adult conceptions of gender-appropriateness. For example, though the data are not current, Best (1983) examined the “gender-role curriculum” that was taught in grades K-6 in US schools. Best reported that by grade 3, through the direct actions of teachers, girls were learning to be helpful and nurturant.

Even in the second grade, boys had learned to expect help from girls in both academic and social domains. Girls stayed inside more during recess and volunteered to help the teacher clean and arrange the room. The boys played outside. By the third grade, the boys had thoroughly internalized the expectations to show strength, to court danger, and not to show affection. In every way, boys were valued to the extent that they were unlike girls (Best, 1983). The agents who modeled the appropriate behavior, and dispensed social rewards and disapprovals, were the peer group of boys. Though the institutional setting was school, for boys at least, fellow students, not adults, were central to learning this unintended curriculum.

Perhaps the most theoretically grounded, socially-based approach to learning and development is the work of Vygotsky (1978). Vygotsky was trained in the tradition of the early Russian associationalists, and though he greatly extended the sense of the reinforcing agent in his approach, his behaviorist grounding rings clearly throughout his work. A good sense of Vygotsky's approach can be gained from the work of his colleague, Luria, who carried out observations and experiments in cognitive development in the early years of Soviet collectivization. Luria thought about the effect of culture on cognitive activity (which he and Vygotsky referred to as sociohistorical shaping) and wondered if exposure to a new culture resulted in simply the acquisition of new habits and knowledge, or whether it resulted in radical reorganization of mental processes (Luria, 1976). This question was clearly rhetorical in nature, and showed the influence on Luria's work of Vygotsky's theme that higher mental processes have their origin in social actions and processes. Vygotsky claimed that individuals are an "aggregate" of internalized social

relationships that have become beneficial to a person's functioning in society. The actual connection between society and individual, between concrete experiences and mental cognition, is mediated through the shared meanings of language and other signs.

Though Vygotsky seems not to have made any direct statements on the possibility of gender differences in socialization, his belief that higher mental processes in children develop through interaction with adults (Wertsch, 1985) can provide the basis for a move to the issue of gender-related differences in ideas about knowledge and learning. For Vygotsky, children begin to use the same forms of behavior in relation to themselves that others initially used in relation to them (Wertsch, 1985). Vygotsky claimed that the effect of culture was not limited simply to the development of habits and acquired information, but extended to basic cognitive structures. This could be seen as at least implicitly related to different cognitive development for females and males if young children are treated differently according to their biological sex.

Cognitive-developmental approaches. Although familial and broader cultural reinforcement is important in the process of forming a personal identity, young children often show behavior that, while not necessarily determined by physiological factors, nevertheless appears to be more intrinsically organized than could be accounted for by an environmental system of differential rewards and punishments. It seems that children have their own ideas about what it means to be a person-in-development, particularly a person with a given gender identity, and that the influence of adults in developing an identity may be less than assumed under a strict social learning approach.

Beal (1994), cited earlier for her observations on social learning, provides several examples of children acting in gender-related ways that are counter to the desires of parents (e.g., the young girl who, despite her mother's wishes to promote non-stereotypical gender development, prefers to vacuum rather than choosing to play with a toy truck). Though adult family members, other adults, and peers no doubt are important reinforcers of expectations and behaviors in general, it may be that gender identity is more the outcome of a selection initiated by the developing child, with the chosen identity then providing the motivation to identify with role models which later selectively reinforce gender-appropriate behavior. Once children gain the knowledge of their sex in a simple definitional way, though they would be free to resist forming conventional ideas about themselves as gendered individuals, a cognitive-developmental approach suggests they would be motivated to learn more about gender role expectations (presumably incorporating dominant cultural cues), to seek out appropriate role models, and to conform to what they think is expected of them as boys or girls by the dominant culture (Beal, 1994).

The idea of a child taking control of her or his gendered development at an early age is not new. Kohlberg (1966) posited that after young children realize the permanence of their sex, they come to actively seek examples of appropriate sex-related behavior. The information gained through observation and subsequent trial is then organized into gender schemas, frameworks that act as cognitive windows through which the world is interpreted (Martin & Halverson, 1981). Such schemas might include knowledge about

what males and females look like, what their typical interests are, and how members of different sexes will act in different situations.

The development of gender schemas are functional in the sense that they greatly simplify the task of deciding what new information is appropriate for incorporation into the cognitive repertoire. Martin and Wood (1987) found that when children were told a story about a little boy who liked to play with dolls, the subjects predicted he would also like to play with a truck, despite clear information to the contrary. New information that is consistent with the expectations generated by the gender identity schemas are more easily remembered than information that is inconsistent with it. And though a child will develop schemas for both sexes, the own-sex schema is more detailed and richer in associations. In children's minds, same-sex and other-sex schemas are not only different, they also are not equal. There is a distinct in-group/out-group phenomenon at work which leads children not only to take on the roles of the like-group peers, but also to carefully avoid any behaviors that characterize the other group (Kuhn, Nash, & Bruken, 1978).

Using social and physical cues from their environment in an active and constructivist manner, rather than simply responding to rewards in a behavioristic way, infants use gross characteristics such as appearance and physiological markers to discriminate between the sexes. As early as the second year, young children begin to show a preference for others of their own sex. It is unclear how this comes about, though researchers have considered the possibility that young children, who understand spoken language before they are able to speak, may be able to construct a proto-gender idea at a

very young age (Beal, 1994). It is also possible that toddlers are able to compare in a proprioceptive way the movement of members of their own and other sex (Bower, 1982). By age 3 1/2, about 80% of the children in one study could correctly identify the sex of other people and knew their own sex (Slaby & Frey, 1975). At this age, judgments are being made primarily on the basis of how people look and what they do than on any understanding of the biological basis of sex. The ideas gained at this age are strong and often take on the characteristic of stereotypes, with abstract qualities such as color, natural objects such as trees and flowers, and artifacts like buildings coming to be associated with one sex or the other (Mullen, 1990).

As children progress through the preschool period they build on their understanding of gender identity to learn that they will remain that sex over time (stability), and finally come to understand that they will stay the same gendered person across situations (consistency). This development has been observed across class boundaries and several cultures (Frey & Ruble, 1992). In fact, once children learn that a person's sex is related to biological differences they become quite essentialist, ascribing differences in behavior to differences in physical characteristics until reaching early adolescence (Smith & Russell, 1984).

Such deeply-held notions of differences in sex and gender-linked behavior may contain ideas about differences in learning and thinking about knowledge. At the least, it seems clear that children are not simply passive recipients of gender-related socialization messages from parents and other adults. Rather, an integral part of gender development appears to consist of active observation of the social environment, with a resulting self-

generated sense of self, accompanied by behavior that is deemed appropriate according to constructed ideas about gender.

Adolescent and Post-Adolescent Cognitive Development

Despite the undeniable developmental importance of childhood (indeed, perhaps because of it), as teenagers we are different from our earlier selves. We continue to change as we move through adolescence and into adulthood. As we grow older our emerging sense of gender is likely to become an important part of our overall identity. Moving through adolescence, most noticeably in school settings, we continue to be engaged in a process of development before coming to see ourselves as possessing a particular gender identity. In addition, several researchers have posited a sequence of change and maturation in the ways we learn and think about the nature of knowledge. These changes are the result of the complex interaction of our physiological selves with the physical and cultural environment. I would argue, as is the case for the development of gender identity, our epistemological beliefs and orientations change over time as well.

In this section, drawing on the work of Perry, Belenky and her colleagues, and Baxter-Magolda in their attempts to account for intellectual development in adolescents and adults, I will describe elements that are treated similarly across all three approaches. The result is a treatment of development that is both descriptive and developmental. The elements have an epistemological focus, containing features associated with the nature of knowledge and the evaluation of truth.

The work of Belenky, Clinchy, Goldberger, and Tarule grew out of the findings of Perry -- and undoubtedly influenced by Gilligan, in many ways it is a reaction to his

findings. The more recent ideas of Baxter-Magolda are themselves an extension and elaboration of those of both Perry and Belenky, et al. All three approaches (even the work of Belenky and her colleagues) emphasize a developmental dimension in their descriptions, which I, in turn, have chosen to represent by three broad categories. They speak to conceptions of knowledge and the nature of truth as part of their approach, ideas which make an important contribution to the development of my research design. The categories are: (1) knowledge as received certainties, (2) knowledge as relative beliefs, and (3) knowledge as conditional constructions.

Though similar in many respects, there are also differences among the approaches of Perry, Belenky, and Baxter-Magolda insofar as the expression of sex and gender-related behavior within categories is concerned. Perry employed only male students in his work. Belenky conducted her interviews only with women. Perry made statements about universal epistemological development without acknowledging his male-only subject pool. Belenky worked exclusively with females and made claims for a pattern of development that was particular to women without considering possible similarities in male development. Baxter-Magolda, on the other hand, worked with both female and male students.

Because there is theoretical and practical overlap among the categories used by Perry, Belenky, and Baxter-Magolda, and because I ultimately want to focus on claims relating to the nature and existence of historical truth-in-knowledge, I will consider each of these approaches as they relate to the three descriptive categories mentioned above. Within these categories, Perry, Belenky, and Baxter-Magolda can be said to consider

knowledge and claims of truth to consist of: (1) received certainties, (2) relative and subjective beliefs, or (3) conditional constructions.

While my claim for the existence of three categories of truth-in-knowledge is based on the largely descriptive work of Perry, Belenky, and Baxter-Magolda, the conceptual processes I assume to undergird the categories are not necessarily the same as theirs. For example, Baxter-Magolda found characteristic male and female expression in each of her categories except the most developed (what I am calling the conditional construction category). I propose to explain these differences in gender expression within categories (and the gendered expression of Perry's male students and Belenky's females) by invoking Bruner's notion of cognitive orientation. That is, I will argue that an individual's narrative or paradigmatic orientation is a pre-existing and stable condition which is expressed as a person moves through the process of developing beliefs about knowledge and truth. More specifically, I will propose that females are more likely to favor a narrative cognitive approach and males a paradigmatic approach to organizing information. If this is the case, it would provide an explanation for the existence of two variations within one nature of knowledge (truth-in-knowledge) category.

In addition, I propose that males are more likely to think of truth-in-knowledge as consisting of received certainties, with females preferring the relative (subjective) sense of truth-in-knowledge. No such gender-linked situation is posited to exist for the development of a sense of truth-in-knowledge as a conditional construction. Thinking of truth-in-knowledge as conditionally constructed is posited to arise from individuals moving beyond earlier gender-specific orientations by means of learning the value of less

avored approaches to thinking about learning and knowledge, and from domain-specific cognitive growth (an increase in so-called content knowledge). As a result, beliefs about knowledge and truth-in-knowledge as conditionally constructed are assumed not to differ in expression across sex and gender.

Knowledge as received certainties. Perry conducted his original research in the 1950's, using male Harvard University undergraduates as subjects. Though his work has been criticized for not including women in the initial theory-building stage, his findings have nevertheless served as the starting point for other, more gender-inclusive research.

In Perry's schema, when knowledge is conceived as received certainties, it is considered to exist "out there" and known by experts (Perry, 1970). Teachers--good examples of experts-- have the obligation to make their knowledge known to students, whose task it is to incorporate the knowledge teachers dispense. If teachers disagree about a specific example of knowledge, one teacher is simply wrong, or perhaps neither has yet perceived the truth. Students operating at this level are not capable of allowing for the possibility that knowledge is created and can be changed once formed.

Belenky and her colleagues (Belenky, Clinchy, Goldberger, & Tarule, 1986) developed a model for describing stages of thinking about knowledge that builds on Perry's ideas about cognitive development. But Belenky et al., using in-depth interviews with a sample of late-adolescent girls and women, extended and supplemented Perry's approach. They identified conceptions about the nature of knowledge which supported claims that thinking about knowledge could differ across gender lines (Belenky, et. al.).

The first three stages of the Belenky model, which show a sense of dualistic or dialectic knowing, fall within the category of knowledge as received certainties.

For example, Belenky and her colleagues described received knowers as those who realize that they can remember and act on the knowledge remembered (received knowers are themselves a developmental step removed from a small number of individuals who are almost literally “silent” with respect to knowing. Silent knowers feel knowledge is possessed only by others; their own condition is marked by a lack of self-perceived ability even to acquire the knowledge of others.) Knowledge for received knowers is clearly external and made known by learned authorities. Authorities, often teachers, have knowledge--friends, while important, only have the proto-knowledge of opinion. Knowledge at this level does not allow for ambiguity. Elements of knowledge -- facts and procedures and conceptual understandings -- are either completely correct or of no use whatsoever. Though Belenky interviewed only women, she stated that women are less likely than men to identify with all-knowing authorities. For women, teachers and other authorities are often seen as unapproachable and not as sources of easily-accessible knowledge.

Although the Epistemological Reflection Model (EMR) developed by Baxter-Magolda contains four stages of intellectual development (Baxter-Magolda, 1992), I believe Baxter-Magolda's categories fit into the three-part model presented here. Baxter-Magolda's model emerged from empirical work carried out over a five year period which chronicled the cognitive changes and development of college students from the time of enrollment through one year following graduation. In addition to delineating overall

differences across categories, Baxter-Magolda found that within a particular category of development, students showed patterns of cognitive activity closely related to gender.

Students at the first stage of the ERM, the stage of absolute knowing, believed that knowledge is objective, external, and certain. The role of the student is simply to acquire knowledge. Within this stage, there are two distinct routes to the acquisition of knowledge. The first is through somewhat private listening or recording of information. The sources of knowledge are the straightforward lectures of teachers and the printed words of texts. Discrepancies in knowledge claims for absolute/private knowers are likely to be resolved by returning to the self for resolution. At this stage, students reserve their voice for building affective support among peers. Baxter-Magolda found more women than men used this absolute/private approach.

In contrast, within the category of absolute knowers a second group can be identified by their more public displays of gaining knowledge. These learners enjoy testing out ideas against other students; they are essentially individualistic learners. They tend to identify with the authority figure who is dispensing the knowledge in question. Among them, the voice of the learner is used to question and clarify. Consistent with Belenky's speculation, Baxter-Magolda found more men than women exhibited this pattern of thinking.

Knowledge as relative beliefs. Following exposure to many examples of conflicting knowledge claims, learners in Perry's model enter a narrow band of authentic relativism. These individuals begin to question the information and ideas presented to them. The defining occurrence for the move into relativism occurs when students

recognize what Perry calls irreducible uncertainty. As a way to get beyond this realization of the existence of non-answers, students began to look for a method to approach knowledge. They begin to learn the rules for navigation within a domain of knowledge, and are eager to transfer a successful approach to knowing across domains (similar to Piaget's notion of horizontal decalage). This is not simply an instrumental activity. Students who begin to adopt a relativistic stance honestly want to make internal sense of the information contained within a domain of knowledge. The agency with relativistic learners has clearly shifted to the student. The relativism at this stage is not a relativism of singular ideas. Rather, it is a non-judgmental accepting of differing approaches to defining knowledge across and within academic disciplines.

In the framework developed by Belenky and her colleagues, the knower who incorporates the ideas of others (the ideas can be expressed in written as well as spoken form) runs the risk of recording conflicting claims for knowledge. As a result, the perceived source of knowledge eventually shifts, moving from outside authorities to the knower herself. Personal experience and the resulting personal constructions of knowledge that arise from such experiences claim a privileged place as justification for claims of knowledge. The result is an approach to knowledge creation that allows for understanding to emerge subjectively from within the individual rather than through the process of logical analysis. At this stage, a person attempting to determine the truth of a proposition is guided more by direct experience and feeling than by the use of more conventional logical criteria.

Similar to the second-level work of Perry and Belenky, the stage of development identified as independent knowing by Baxter-Magolda fits within the conception of knowledge as consisting of relative and subjective beliefs. Independent knowing is marked by the widespread belief that most knowledge is unclear. Students who are independent knowers operate with a core assumption of uncertainty. Thinking for oneself is valued, with much importance given to determining the best process of making sense of conflicting knowledge claims. Learning the process of careful thinking provides the grounding for further intellectual work. Independent thinkers are open to interpretations of peers and teachers, though their decisions are made in relative isolation from others.

As is true for other levels in Baxter-Magolda's model (with the exception of the highest level), there are two gender-linked categories of independent learners. Women are more likely to show an inter-individual pattern of learning. Inter-individual learners actively seek situations in which the opinions of others are expressed. The voice of the learner is strong, and inter-individual learners are not reluctant to take public stands in support of their positions. At this level, a connection between school and personal lives is developed. Information from peers as well as teachers is taken in, examined, and compared with one's own ideas. For the first time, a condition of equal worth as a participant in the process--of mutual respect that allows for the integrity of both self and other--exists for the learner. A student operating at the independent level feels herself to be a fully functioning member of the intellectual community. As the learner's voice becomes more sure, students showing an inter-individual pattern of learning become closer to the authority figures from whom they were formerly more distant.

Men who function at the independent way of knowing, on the other hand, are more likely to exhibit what Baxter-Magolda calls an individual-pattern in their intellectual activity. Individual-pattern learners genuinely know the value of listening to the views of others, but they continue to focus on the acceptability and provability of their own views. Despite their feeling that knowledge is uncertain, confidence in their ability to find knowledge leads them to the verge of separating from the views of traditional authority figures. Independent learners showing an individual pattern also move for the first time from an identification with teachers to a serious consideration of the opinions of peers.

Knowledge as conditional constructions. The latter stages of Perry's model represent a broad category with the development of conditional commitment to one representation of knowledge as the salient feature. Students who exhibit conditional commitment recognize the tentative nature of their commitment, but are motivated by a sense that knowledge claims can be accepted as if they have been satisfactorily justified and have met at least provisional conditions of truth. Along with this, there is a sense at the stage of conditional commitment that the individual needs to become active in creating knowledge. To participate in the process of generating knowledge, students need to believe that the results of trying to construct knowledge (even if the outcome of the activity is considered to be conditional) will be more positive than remaining in the state of relativism, where no claim can be judged more worthy than another.

Perry's learner is now able to recognize the value of commitment in the face of ultimate uncertainty, especially when this is modeled by an older teacher. Part of being committed is to accept the paradox that a person may hold strongly to a belief, while

knowing that the underlying question or task may be unsolvable. To show commitment means to seek out the ideas of others and accept the risk that their ideas can affect the strength of one's commitment.

The model developed in Women's Ways of Knowing concludes with a position of integration labeled constructed knowledge. The emerging constructed knowledge knower senses the limitation of considering knowledge as based either solely on personal intuitive constructions or more external, authoritative, rational constructions. An individual at the constructed knowledge position seeks to integrate the processes of reason and intuition, and works to find a place for an interpretation of the experiences of others. The shift from procedural (either separate or connected) to constructed knower is marked by the posing of essential questions about the self and the creation of knowledge. The nature of identity becomes important. A narrative sense of the self is constructed; there is an attempt to integrate the past and an idea of a future with the present. At the same time, more tolerance is shown for ambiguity and internal contradictions. Through the process of forming a sense of self and looking for a unique position from which to speak, a person functioning at this stage comes to believe that knowledge is the result of someone's or some group's conscious construction. From this position, it is a short step to believing that the ultimate constructor of knowledge (and hence understanding) is the learner herself.

Knowledge for constructed knowers becomes heavily dependent on the context in which it is created. The outside culture and internal frame of reference of the knower will interact to produce a setting in which questions are asked and answered. Theories become

not containers and keepers of truth, but time and culture-bound models for interpretations of experience. At this stage, knowing involves a consideration of the assumptions and conditions within which the problem is cast. There is often a normative consideration of the consequences of one's actions, with a desire to do the best possible for all involved in the process.

Contextual knowing, the fourth level of the Baxter-Magolda's Epistemological Reflection Model, shows the same qualities of synthesis and lack of certainty as the approaches of Perry and Belenky. The essential element at this stage of development is the integration of one's own ideas and the knowledge of others. At the contextual level it becomes necessary to incorporate claims generated by others in the process of reaching a conclusion on any given question of what counts as knowledge. This stage is also marked by the first occurrence of the use of judgment in regard to which claims of knowledge generated by others will be accepted and which will be discounted.

The contextual learner continues to view knowledge as less than fully formed and revealed, but the lack of certainty exhibited by a student at this stage is different from the relativism that is part of earlier stages. The richer context available for college students who are in advanced stages of study (e.g., internships and other project-based approaches) offers the framework for judging claims and conflicting opinions. Students develop what Baxter-Magolda calls deep-structure learning at the contextual stage, through real-world applications and feedback to their earlier, more conceptual learning.

Baxter-Magolda did not identify any gender-related patterns at this level. She hypothesized that this may be due to a converging of earlier differences between men and

women, with men complementing their mastery and individual ways of thinking about knowledge with greater connections to the legitimate views of others, and women becoming more willing to become judgmental while maintaining their fundamental interpersonal and relational framework. It may also be that the process of developing deeper understanding in advanced learning tasks requires a student to draw on both individual and inter-individual strategies. Or, as I claim on behalf of Bruner (see below), it may be that both paradigmatic and narrative cognitive orientations are called into play in the process of gaining greater content knowledge at the level of contextual learning.

In sum, Perry, Belenky, and Baxter-Magolda have each made unique contributions to understanding how we think about the nature of knowledge and claims for truth. Yet it is also clear that, taken collectively, their work shows a marked degree of convergence which can be represented by a three-stage model with overtones of gendered development at each stage. Though not developmental in a strict Piagetian sense (Women's Ways of Knowing makes no claim for a disequilibrium mechanism that initiates development, for example), and making no claim to examine the value assumptions that probably underlie a developmental paradigm, each approach is developmental in at least a loose sense, containing the idea that constructed knowledge is more adequate and complex than the consideration of knowledge as received certainties (see Goldberger, 1996, who claimed the five WWK positions can be arranged from less to more adequate, so long as culture and context are taken into account).

These approaches also contain a similar loose sense of the nature of truth. Though Perry and Belenky did not make direct mention of what truth looks like to different

knowers, Baxter-Magolda did. Her descriptions of the nature of knowledge at each level are essentially what I am calling truth-in-knowledge (e.g., the nature of knowledge at the independent knowing state is uncertain -- everyone has their own beliefs). These are not traditional senses of truth such as correspondence or coherence theories of truth, but rather workaday conceptions about the origin and worth and durability of the knowledge claims (in this case, claims for historical knowledge) that go before them. Considered in this way, ideas about truth (or truth-in-knowledge) by definition are linked with the category of knowledge that creates them. It is not conceptually possible to think of knowledge as absolute and unchanging while simultaneously believing that the nature of truth-in-knowledge is to be judged, for example, on the basis of evidence in context.

The Work of Gilligan and Bruner

The literature review to this point has shown how children develop a sense of themselves as gendered individuals, and how they bring this sense with them as they enter adolescence. Furthermore, the work of Perry, Belenky, et al., and Baxter-Magolda argues that individuals approach learning and thinking about knowledge in what can at least be loosely called a developmental fashion: first believing knowledge to be what has been discovered or constructed by others; later moving through stages of uncertainty marked by relativism and subjectivism; and emerging at some level of tentative commitment grounded in a mix of knowledge claims created by oneself as well as the claims of others.

Perry, Belenky, Baxter-Magolda, and others who have written about developmental issues have contributed much to our understanding of the processes of intellectual growth. Yet, despite these contributions, I would argue there is a need to

think about knowledge in another way as well -- a way that moves beyond beliefs about the development of knowledge and truth claims to the place where we organize the information that is taken in and processed in the task of creating meaning. I believe my research will show this cognitive process, concerned with the organization of knowledge, differs across sex and gender. To provide the background for this assertion, I move to the ideas of Gilligan and Bruner.

Gilligan and moral thinking. The work leading to Gilligan's claims about general cognitive and moral development took place over twenty years ago. Though her treatment of moral development particularly has been criticized almost since it first appeared (Cherry, 1994; Flanagan & Jackson, 1987; Friedman, Robinson, & Friedman, 1987; Garmon, Basinger, Gregg, & Gibbs, 1996; Hepburn, 1993; Keefer, 1993; Stiller & Forrest, 1990; Vreeke, 1991), Gilligan's call for the recognition of cognitive development as a gendered process nevertheless allows for the move to the possibility of sex and gender differences in both the conception of the nature of historical truth and cognitive orientation.

Early on, Gilligan questioned Kohlberg's assumption that women were somehow less developed cognitively because they did not demonstrate the level of preference for independent thought and action that was observed for men. Through her own research, employing a methodology that used lengthy and probing interviews, she found that women could differ from men in respect to their approach to defining and working with matters of morality. Many men in Kohlberg's studies tended to define responsibility in terms of limiting what would otherwise be unchecked action, whereas the women in Gilligan's

studies were reported to characterize responsibility more contextually, infusing it with a sense of connection with others (Gilligan, 1982/1993). Undoubtedly influenced by the work of Chodorow (as cited in Gilligan, 1982/1993), Gilligan reported that the women in her studies, thinking about responsibility, started from positions of connectedness with others, only later coming to see the worth of tempering such an approach with a more independent and self-centered position.

Gilligan's work allows for the concept of historical truth to be considered in a gendered context (are females more likely to think of historical truth in relative terms than males?). In addition, her work with moral decision-making also provides the context for considering cognitive orientation in a gendered manner. Gilligan contends that women begin by defining the problem of moral decision making in terms of conflicting responsibilities; men, from consideration of competing rights. Resolution of moral dilemmas for women comes from thinking that is contextual and narrative. For men, thinking is formal and abstract. Gilligan reported that, in making moral decisions, women were more likely to request missing information about the nature of the actors and their specific actions (Gilligan, 1982/1993). This was in contrast to men, who were assumed to engage in a hierarchical ordering of principles in their moral decision-making activity. By way of explication, Gilligan invoked the operative metaphor of the web. Women work to extend and to function from within webs of relationships; men presumably spin as well, but the result of their construction is apart from themselves, hierarchical and objective.

Bruner and modes of cognition. In contrast to Gilligan's focus on gendered differences, Bruner can be considered more of a general systems theorist. As noted

earlier, Bruner has argued for the existence of what he characterized as two “irreducible” modes of cognitive functioning (Bruner, 1985; 1986) which are active as we engage in any cognitive activity. These two distinct cognitive modes--paradigmatic and narrative in Bruner’s terminology--serve to organize and reorganize all internal and external human perceptions. Paradigmatic thinking is informed by the ideal of creating mathematical or otherwise formal systems of description and explanation. Narrative thought is temporal in nature, and is useful insofar as it produces what Bruner calls “believable” accounts of the world.

Bruner has not offered any opinions regarding differences across gender in the use of paradigmatic or narrative thinking (he eschewed any interest in what he called “individual differences” within the overall model; personal communication, January 21, 1996). Nevertheless, I believe his conception of cognitive structure, when considered in terms of Gilligan’s conceptions of thinking in the moral domain, offers an approach to considering gendered differences in organizing information and knowledge. I see his generalized paradigmatic and narrative modes as mapping onto the justice (formal and abstract) and care (contextual and narrative) dimensions of Gilligan’s conception of moral decision-making. In fact, though not examined empirically in any depth in this study, Bruner’s conception of a characteristic manner of organizing the perceptual world in either narrative or paradigmatic ways may serve to explain the differences between females and males within categories of epistemic development hinted at by Perry and Belenky, and seen directly by Baxter-Magolda. I referred to this possibility before when I said that Bruner’s ideas of narrative and paradigmatic thinking could be used to account for

gendered expressions of behavior within Baxter-Magolda's nature of knowledge -- and what I call truth-in-knowledge -- categories. For this reason, I would like to consider Bruner's thoughts on modes of cognition more fully.

According to Bruner, paradigmatic thinking involves consideration of general causes. Principles of inductive reasoning are used in paradigmatic thinking to test for truth. Paradigmatic thinking deals with the empirical world and with unsensed worlds that are logically possible to know. It favors denotation over connotation. Thinking in the paradigmatic mode means thinking about events of the real world (the social as well as the natural world), and then considering a level of abstraction that gives the unequivocal description of these events a larger order, structure, and more inclusive meaning. Paradigmatic speaking commits a speaker to a literal description of the world. A writer thinking paradigmatically means, by definition, exactly what she says.

The narrative mode considers matters "close up". Narrative thinking actually constructs two landscapes at the same time. One is an action landscape, the actual carrying out of the story at hand. The other is the landscape of the consciousness, how the actors feel and think. Joining the landscapes, narration becomes the process of being concerned with the explication of human intentions in the context of action (Bruner, 1985).

Narrative thinking does not represent an early, incomplete, or inferior form of paradigmatic thought. Although the ability to think in terms of observable action (in literary terms, following the story line) develops early in childhood, it is a much greater cognitive challenge to take on the point of view of the actors in the narrative.

Understanding both intention and action does not occur until later in development.

Narrative thinking attends to events in the physical or social world and the reaction of people to them. It is a discourse that through the “relative indeterminacy” of a text, lets the speaker represent more than is actually said. Narrative allows (perhaps requires) the listener or reader to initiate “performances of meaning” rather than receive a prepackaged and complete meaning. Through the use of linguistic “triggers”, an author enriches the narrative with a presuppositional grounding in order to “pack” the text with material which will be fashioned into reader-constructed meaning (Bruner, 1986).

Again, it should be noted that Bruner makes no claims for the superiority of one mode over the other. His categories are nominal. Each arises “spontaneously into existence in the functioning of human beings” (Bruner, 1985, p. 103). Insofar as we can think of both modes as being made up of component parts (Bruner calls them “constituent processes”), even when the parts share the same label, they have different properties (e.g., both modes of thought use the idea of contradiction, but it has an almost logical-operator use in paradigmatic thinking; in narrative thought it is simply descriptive, having no controlling character).

Though Bruner made no reference to possible differences across gender in the use or preference for using the two modes of thinking, he does offer a possible way into thinking about a relationship between the mode of cognition and academic performance in schools and on tests. Describing strategies employed for making inferences in research settings, Bruner (1985) noted that different contexts for judgment may activate different representations of the problem and the use of different inferential strategies to solve it.

For Bruner, the different strategies arise from the two different modes of thought. And though these observations were made in the context of laboratory research, they may point the way to a connection between ways of organizing knowledge and differences in performance in an academic setting. I believe this line of thinking has specific relevance for higher-level learning involving constructed knowledge, in which a student may need to draw on the non-preferred mode of cognitive organization (narrative or paradigmatic) in the process of creating meaning and understanding.

It may be that men do not think of knowledge as an “idealized external form”. It may be that women do not operate from a “biblical conception” of knowledge, believing it emerges from human relationship (Gilligan, 1982/1993). There may not be an underlying difference in the ways women and men construct their worlds. Nevertheless, I believe there are gendered relationships between and among the ways a person thinks about the nature of knowledge (particularly the conception of truth-in-knowledge), how they organize information in the process of creating meaning, and their subsequent academic performance. Accordingly, I propose to develop instruments that will operationalize the two conceptions discussed above (truth-in-knowledge and cognitive orientation), and determine whether they are related to academic performance in a particular academic domain.

Though I believe these relationships among epistemic beliefs, cognitive orientation, and achievement will show distinctive patterns across sex and gender, I also acknowledge -- invoking the sense of parsimony -- that differences in academic performance may simply be reflections of deeper differences in general or domain-specific ability between males and

females. In addition, there may be other plausible explanations for gender-related performance differences across the different formats used in tests of achievement. Both these possibilities need to be given the opportunity to explain differences in achievement between females and males. I turn to these two matters below.

Achievement and Ability

There is no theoretical reason why ability and achievement should be confused. In everyday use, ability refers to a state of relative preparedness, often with the connotation of being present before other training or development. With respect to academic matters, ability is an input variable. Achievement, on the other hand, is an outcome. It is the result of many factors, ability being one of them. Yet despite the conceptual clarity with which we can distinguish them, the literature is confusing on matters of achievement and ability. For example, research into the relationship between gender and achievement has revealed less about achievement than about the related, but conceptually different, issue of gender and cognitive ability. And though it is not difficult to argue for the existence of a connection between achievement and ability, it may be that how achievement is assessed may favor one gender independent of underlying differences in ability (indeed, the possibility of such a state of affairs is an important part of my research hypotheses).

But to return to the point, the literature cited in this section is almost exclusively drawn from the work of researchers who were looking at similarities and differences in underlying cognitive abilities across gender. It should also be noted that much of the evidence used to make claims about cognitive ability are themselves based on differences in performance in academic settings - in short, achievement. At the least, we need to be

careful about how the two terms are used. I propose to move from thinking that both ability and achievement refer to successful (academic) performance, and instead (as stated above) to think of ability as one of several factors that contribute to achievement.

There has been an interest in what Fairweather (1976, p. 231) called the “psychological differentiation of the sexes” since at least early in this century. By the end of the 1920s, a conventional wisdom had emerged that females possessed superior verbal abilities and males had better quantitative and spatial skills (Allen, as cited in Fairweather, 1976). Since that time, psychologists have examined many non-cognitive differences between males and females (e.g., reaction times and motor movement; the behavioral concept of aggression) while continuing their work on ability and achievement in the cognitive domain.

The body of research on the relationship of sex and cognitive ability, nearly fifty years in the making, was itself evaluated by Maccoby and Jacklin (1975). Maccoby and Jacklin undertook a comprehensive narrative review of studies that were conducted and published through the early 1970’s. They concluded that at the start of adolescence the cognitive abilities of the sexes begin to diverge, with females moving ahead of males in verbal ability throughout high school and possibly beyond. The opposite situation was found for mathematics and spatial achievement, with the research showing smaller, but nevertheless consistent, results favoring males.

Almost immediately, Maccoby and Jacklin’s work was criticized. Fairweather (1976) questioned the study on several methodological points, noting that Maccoby and Jacklin gave equal weight to what he called good and bad studies. According to

Fairweather, one finding showing differences was likely to be given as much attention as a number of negative findings in follow-up studies. Fairweather concluded that there were few sex-related differences in cognition and recommended against further research in the area.

Hyde (1981) took issue with Maccoby and Jacklin's conclusion that gender differences in verbal ability, quantitative ability, and visual-spatial ability were "well established". Hyde re-examined the data reviewed by Maccoby and Jacklin using a meta-analytic approach which was more quantitative than the narration and simple enumeration employed by Maccoby and Jacklin. Hyde included all of the Maccoby and Jacklin studies which were done in the US and carried out with adolescent and older subjects. The median values of the effect sizes (reported in difference units, the numerical difference expressed in standard deviation units) for verbal, quantitative, and spatial abilities ranged from .24 to .45. According to Hyde, sex differences appeared to account for no more than 1-5% of the population variance in all three areas. While allowing that such differences can still result in sizable differences in the tails of distributions (with a .4 d between populations, there would be twice as many representatives of one population above the 95th percentile) Hyde felt the overlap between populations was too great to consider the sexes as differing in abilities in an overall meaningful way.

Hyde and Linn (1988) also carried out a meta-analysis of research into gender differences in verbal ability which was done after the work reviewed by Maccoby and Jacklin. The overall d in this meta-analysis was small and favored females in each age group studied (though when a weighting for study size was included, there was a shift to a

slight favoring of males). The observed differences were smaller for the most recent studies reviewed, a trend similar to that noted by Chipman, Brush, and Wilson in their report on gender differences in performance on assessments in mathematics (as cited in Hyde and Linn, 1988). Hyde and Linn concluded there was no evidence to support a belief in a difference across genders in either generalized verbal ability or within verbal subtypes.

An additional point needs to be made about the findings of meta-analytic studies of differences in cognitive abilities. It is difficult to know whether the studies included for review were measures of ability or indicators of achievement - or some mix of the two (e.g., a measure of ability may be confounded by the selection for study of students who were high achievers). Maccoby and Jacklin speak about ability, but their conclusions are likely to be at least partially influenced by studies which did not control for achievement when looking for differences in ability.

Though the meta-analytic work of Hyde (and the later work of Hyde and Linn) presents a strong case for the absence of sex differences in cognitive ability, it should be remembered that in her 1981 review of the earlier Maccoby and Jacklin research, Hyde found effect sizes favoring males in visual-spatial ability nearly twice those favoring females in the verbal area. Halpern (1992) has commented on these differences, pointing out that sex-differences in certain quantitative sub-domains (e.g., problem-solving and mental rotation of three-dimensional figures) continue to show nontrivial effect-size differences in populations of older adolescents and adults (though this could represent a

statistical artifact which follows from the unequal overrepresentation of males in the top tail of the achievement distribution in the general quantitative area).

In another commentary, Eagly (1995) claimed that sex differences in cognitive achievement, though modest, are nevertheless within the range of the small-to-moderate differences that mark other psychological research. In addition, such sex-difference findings do not appear to be any more inconsistent across studies than cross study findings in other areas of psychological research. Eagly concluded that the common description of empirical research on cognitive differences as showing small and unstable and artifactual cognitive differences across studies, has arisen “in part from a feminist commitment to gender similarity as a route to political equality” (Eagly, 1995, p. 149).

In any case, ability differences need not be seen as inherent conditions that preclude the possibility of equity in subsequent achievement. Furthermore, with the large overlap in populations even in cases where meaningful effect-size differences are found, it would seem that cognitive abilities are distributed in wide measure among both males and females. After all, populations that show differences in the moderate .2 to .5 effect-size range will still overlap by roughly two-thirds.

Overall, the research on sex and cognitive abilities presents a mixed picture of generally small to moderate differences between females and males. Perhaps more importantly, studies that have considered what Eagly (1995) has called the more demanding question of why the sexes sometimes seem to differ considerably in ability and other times only moderately or minimally have produced an inconclusive sense of the origin of the differences and what the future course of research into this area should be.

Looking more directly at achievement, in studies of performance of high school students a number of different agents and conditions have been hypothesized as interacting with sex and gender to produce different academic outcomes. There has been research investigating the role of teachers (e.g., attribution theory) and the importance of the classroom learning environment (Bennett & Bennett, 1994; Payne, 1992). Other studies have examined the effect of parents (parenting style and orientation toward academic achievement) and family configuration and family income (Feldman & Rafferty, 1993; Mulkey & Morton, 1991; Paulson, 1992, 1994; Smith, 1992). Psychological factors such as response to success or failure, academic and non-academic self concept, cognitive style, locus of control (internal/external), and orientation toward learning have been investigated for their association with achievement (Hadfield & Maddux, 1988; Moody & Gifford, 1987; Reap & Cavallo, 1992; Roberts, 1986; Robison-Awana, Kehle, & Jenson, 1986; Stipek & Gralinski, 1991; Williams, 1993). There are claims that the structure of general-level knowledge (e.g., abstract conceptualization) or discipline-level and specific skill-oriented content affects achievement, at times in an interactive way with sex and gender (Agruso, 1990; Doolittle, 1986; Hahn, 1982).

Since it is the nature of research to report findings that show differences, it is not surprising that many of the studies cited above note differences in achievement across sex and gender. But overall, the findings need to be viewed cautiously. In many cases, course grades and GPA were used as an indicator of achievement. And while grades are certainly not an unreasonable choice as a marker for academic achievement, it may be less reasonable to assume that the variable being investigated is the only determinant of course

grades or GPAs. In a number of studies, researchers acknowledged the probable interaction of variables (e.g., the need to study motivation in content-specific areas), but did not predict the effect of the interaction on achievement. In any case, there is little to be gained from this research as to the direction of causality, whether any one ability factor is affecting achievement, or if the social and psychological determinants of achievement are contributing to the development of more developmentally permanent ability.

Research into the explication of differences in cognitive abilities can be represented by two competing sets of approaches - the first emerging from social-psychological theories of behavior, the second from biological and developmental approaches. The social-psychological based explanations ascribe an important role to social roles, feature gender-based expectancies about behavior, and consider differences to arise constructively as a byproduct of social interaction. Biological and developmental theories, on the other hand, treat sex differences as arising mainly from qualities that have a physiological basis for their expression. At this time it is not possible to know the influence of each of these different sets of hypotheses on ability and subsequent achievement.

To conclude the discussion of whether and why sex-related differences in cognitive ability and subsequent achievement do or do not exist, I turn again to the work of Halpern (1992). Halpern presented examples of research data on cognitive differences in ability that cannot be explained by social-psychological hypotheses. She also presented data that cannot be explained by biological hypotheses -- and finished with a consideration of data unexplained by either type of hypothesis.

The data that most strongly support a biological contribution to differences in cognitive abilities is found in research on spatial and verbal abilities in which sex is considered along with handedness. Left-handed females perform better on tests of spatial abilities than right-handed females. A reverse pattern has been found for verbal abilities. Right-handed males do better than left-handed males on spatial abilities measures, but left-handed males were better on tests of verbal abilities. Halpern considered these findings convincing evidence for the role of cerebral lateralization as a contributing factor in producing cognitive sex differences (Halpern, 1992).

However, Halpern also cited several examples of large changes in educational and professional achievements of females (and to a lesser extent, males) in the past 30 years that support social-psychological interpretations of ability differences. The finding that females receive higher grades in school (clearly a measure of achievement) even in areas in which males perform better on ability tests, shows that a large portion of the sex differences effect is social-psychological in origin (Halpern, 1992).

Finally, Halpern offered examples of data from research into differences that are not explained by either a biological or social-psychological approach. Girls performed better on some types of mathematics tests (e.g., computation) and boys on others (e.g., numerical problems). Boys are better on some kinds of verbal tasks (e.g., verbal analogies) and girls on others (e.g., fluency). Neither theory can account for changes in cognitive abilities over time (e.g., females have lost their long-time advantage on the SAT/V in recent years). Sex differences favoring males are most frequently found in

timed tests. Such findings could be attributable to social-psychological explanations or as evidence of sex-based biological differences in the central nervous system (Halpern, 1992).

The research on cognitive ability and achievement -- and the extensive review and discussion of this research -- shows convincingly that questions regarding the existence and causes of differences in ability and achievement across sex and gender are complex. It is also certain that many questions remain unanswered. Considered in the context of my study, the research might be taken to suggest a slight advantage for female test-takers on constructed-response parts of examinations (e.g., essay questions and document-based questions). On the other hand, there would not seem to be any underlying cognitive mechanism to explain the superior performance of males on test items presented in a multiple-choice format (though there may be contributing social-psychological factors at work - see Multiple-Choice Testing and Gender below). In any case, it is clear that the existing research on gender, and on ability and achievement, cannot be taken as providing an explanation of why there is such a large difference in performance between females and males on different parts of the AP US History examination.

Multiple-Choice Testing and Gender

Until the early 1980's, research investigating differences in achievement between females and males did not focus on the mode of assessment. For example, the work of Maccoby and Jacklin (1974) and the critique of Maccoby and Jacklin by Hyde (1981) were concerned with questions about the existence and meaningfulness of differences in types of cognitive ability across sex (which was most often inferred from

differences in achievement). No attention was given to the formats of the item types employed in the studies reviewed by these researchers.

However, changes in the development and delivery of standardized testing in the British secondary system, most notably the addition of multiple-choice items to examinations that formerly consisted of questions presented only in essay format, has led in Great Britain to a concern with matters of test bias across racial, ethnic, and gender groups. Much of the work into differences in performance across item-types in the last 15 years arises from this British psychometric interest.

Though researchers have commented on the possible influence of item-type on performance for some time (see Bolger and Kellaghan, 1990), the earlier studies did not consider whether this influence was associated with sex or gender. The earliest work in which item-type and the sex of the examinee were considered appears to be Murphy (1982), in which he examined scores on 16 General Certificate of Education (GCE) examinations from 1976-1979 in England. Following the introduction of an objective portion to these examinations in 1977, males obtained 10% more As, Bs, and Cs (at the O level) than female candidates. Before the introduction of the objective section, the distribution of these grades had been very similar for females and males across a wide range of subjects. Murphy attributed the differences in performance to general factors related to the format of the tests rather than to the specific items contained within them.

Bolger and Kellaghan (1990) looked at scores for males and females on three examinations (English, Irish, and Mathematics) taken by students in Ireland at age 16 that contained both multiple-choice and constructed-response items. The content tested was

the same for both item types. Males scored higher than females on every multiple-choice measure, including English. The reverse was true (favoring females) for scores on constructed-response items. The effect sizes were small (between .10 and .20) but consistent for all tests.

In the United States, Ferber, Birnbaum, and Green (1983) examined the achievement of females and males in a college introductory-level course in Economics. Ferber and his colleagues constructed an examination composed of brief essay questions and 15 multiple-choice questions which measured the same conceptual material. The test was given as part of the final examination for the course. The results of the study showed that males performed significantly better on the multiple-choice questions and slightly better on the essays than females.

The examinations of the College Board's Advanced Placement Program are the only tests given in large numbers across the United States that contain items in both a multiple-choice and constructed-response (essays, problem-sets, document-based questions, etc.) format. As a result, much of the research involving performance across gender on item types (including multiple-choice items) has originated at the Educational Testing Service (ETS constructs AP examinations for the College Board). Mazzeo, Schmitt, and Bleistein (1992) examined scores and patterns of responses for males and females on four large-volume AP examinations given in 1986 and 1987 (Biology, Chemistry, English Language, and US History). They found differences between males and females to be approximately zero (expressed in standard deviation units) on the constructed-response sections of the four examinations. For the multiple-choice parts of

the tests, the differences were on the order of .33 standard deviation units and favored males. These findings, along with those of Ferber, cited above, were consistent with several other studies in which females did as well as or better than males on constructed-response sections of standardized examinations, with males performing better than females on questions presented in a multiple-choice format (Traub & MacRury; Klein; Breland & Griswold, as cited in Mazzeo, Schmidt, & Bleistein, 1993).

The literature shows that males achieve at higher levels than females on items presented in a multiple-choice format. On the other hand, females--even when they do not receive higher scores than males on the constructed-response questions -- appear to perform better in a relative sense on essay and other constructed-response type questions (the score on the document-based question contributes more to the final score on the AP US History examination for females than for males). Many possible reasons for the differences between male and female performance on tests using multiple-choice items have been proposed. Murphy (1982) was circumspect regarding male superiority on multiple-choice questions. For tests in domains other than mathematics, he proposed that the superior performance of males appears to be caused by "general factors" related to the format of the tests. In mathematics, the emphasis on numerical problem solving was given as the most likely reason for male dominance. Murphy speculated (without citing any empirical support) that multiple-choice questions may require specialized problem-solving skills which are unevenly distributed (presumably by gender) among the student population.

Bolger and Kellaghan (1990) claimed girls' superior verbal ability was the reason for female advantage on constructed-response questions and offered (again, as ex post facto speculation only) a social-psychological explanation for boys' superiority on multiple-choice type questions. They hypothesized that girls may respond less well to novel situations, and to the extent that multiple-choice questions assess knowledge that isn't directly connected with the taught curriculum, girls may perform more poorly than boys. Such an explanation seems to follow the assumption that boys have better developed test-taking/problem-solving skills than girls, an assumption for which Bolger and Kellaghan did not offer any support.

Orientation toward an academic discipline may interact with gender in determining performance on multiple-choice questions. Bell and Hay (1987) examined performance on a test of English language, one of five Tertiary Admission Examinations (TAEs) taken at the end of year 12 by students in Western Australia as part of the university admission process. The test contained multiple-choice and constructed response items. Bell and Hay found an interaction between the discipline orientation (students were classed as either verbally or quantitatively oriented according to the number of examinations taken from two groups of subjects) and the sex of students with performance on the test. Females scored higher on all parts of the test, but the difference was least on the multiple-choice part of the test and the difference in performance across sex was much less than differences between students classified as either arts (humanities) or science oriented.

In their investigation of differences in performance on the AP US History test, Mazzeo, Schmidt, and Bleistein (1992) looked for multiple-choice items which may have functioned differentially for males and females who performed at similar levels on the test as a whole. They found only two items (from a total of 100) that met the criterion of differential functioning. After excluding these items and rescoring the examination, they found only a small change in the standardized difference in performance between males and females. Male performance on multiple-choice questions was still clearly better than female performance. Mazzeo, et al. also found very little change in differences between females and males when both multiple-choice and free-response scores were adjusted for reliability. But Mazzeo also recognized that females performed better on items presented in a constructed-response format, acknowledging that the relatively better performance of females on constructed-response tests may be due to differences in overall sex-specific levels of mastery in the competencies required for good performance on constructed-response item types (the particular competencies were not defined, but they are likely to resemble those listed by Breland, Danos, Kubota, and Sudlow, noted in The Assessment of Achievement in History, below).

Several researchers have examined the role of guessing behavior as it relates to items presented in a multiple-choice format. Differences in guessing patterns for females and males on multiple-choice items were first noted in 1971 in work of Harris (as cited in Bolger & Kellaghan, 1990). Harris, in his work with adolescents, found that girls showed less propensity to guess -- they were more likely than boys to omit items they felt unsure

about. When required to guess on items originally omitted, girls showed a greater gain than boys.

If boys act on their hunches more often than girls when they feel at least partially confident of their knowledge, male multiple-choice test-takers would almost by definition receive higher scores than females. Of course, unless we accept the idea that successful performance on multiple-choice questions requires a special set of problem-solving skills (consistent with Murphy's line of speculation noted earlier), then such observed differences in guessing behavior, while interesting, give no support for an explanation of why the behavior occurs.

Ben-Shakar and Sinai (1991) examined male and female guessing and proposed that differences in approaches to responding to questions posed in a multiple-choice format may account for at least some of the difference in performance. They argued that responding to items when not certain of the answer (i.e., guessing) is a complex activity. Guessing can be an act of random selection or a considered activity informed by partial knowledge. It may also depend on situational factors like test instructions, time pressure, and the perceived difficulty of the items. Ben-Shakar and Sinai hypothesized that boys may show a greater degree of risk-taking on academic tasks such as test-taking. They studied performance of a large group of 9th grade students and an older group of students taking a college-entrance examination. There was no penalty for guessing on either set of tests. Using the Ziller statistic (essentially a ratio of wrong answers to the total wrong answers and omitted items), Ben-Shakar and Sinai found that even on the subtests that showed female superiority, males were less inclined to omit items. Even permissive

instructions regarding guessing did not eliminate general differences in the tendency to guess. When a correction for guessing was factored in (changing the scoring from a counting of right answers to assessing a penalty for wrong responses), females benefited more than males on all subtests in both the high school and college-entrance batteries.

Atkins, Leder, O'Halloran, Pollard, and Taylor (1991) found a similar pattern of guessing in a study of performance on the Australian Mathematics Competition, a 30 question, multiple-choice examination taken by students in grades 7 through 12 in Australia. Boys were found to produce slightly higher Ziller statistics at all grade levels, though the difference appeared to decrease for older students.

Hassman and Hunt (1994), in their work with guessing activity on multiple-choice questions, concluded that students often have at least some knowledge of the question. They proposed that possessing incomplete or partial knowledge favors males because they are more likely to be "test wise" than females. For items presented in a multiple-choice format, test-wiseness refers to the ability to respond advantageously to extraneous clues, and therefore, to obtain credit without certain knowledge of the subject-matter being tested (Hassman and Hunt, 1994). The existence of test-wise ability must remain speculation -- Hassman and Hunt did not provide an explication of the factors that could be considered part of a general test-wise ability, nor did they refer to any empirical testing of how their test-wise hypothesis interacted with partial knowledge to produce better scores for males than females.

Despite the questionable grounding for their contention of the existence of a generalized test-wiseness, Hassman and Hunt carried out other research which examined

sex differences in responses to questions presented in multiple-choice formats. They argued that males and females are likely to differ in their perception of correctness of the answer they select on multiple-choice format questions. To test this hypothesis, Hunt (as cited in Hassman and Hunt, 1994) developed a method to assess the sureness of an answer (defined as the student-perceived correctness of each answer). Using Hunt's instrument to assess sureness, students gained or lost points depending on both the correctness of the answer and the accuracy of the self-assessment (in the Hunt methodology, if a person felt strongly that she did not know the correct answer to the question and indicated that strong "wrongness" feeling, she would still receive some points toward her total score on the test). On a 50-item test composed of old SAT items (verbal and quantitative items in a ratio of 4 to 1), females who were asked to indicate their perceived correctness of answers outperformed females who took the test without indicating sureness. Hassmen and Hunt claimed that the act of asking for self-reflection caused females to undertake a thoughtful review of their knowledge base (though one could also argue that it was perceived as a call for closure for students who might otherwise be considering many connections among and between the question and possible answers).

Looking only at the scores earned on the sureness scale, Hassmen and Hunt found no significant difference between the scores of males and females. The percent "sure and correct" did differ, favoring males. The "unsure but correct" score was higher for females as was the score for "unsure but wrong", which Hassmen and Hunt took to indicate a more uninformed condition regarding the topics on the test.

Though there are differences across disciplines and areas of study, the research on sex and test performance supports the claim that males perform better than females on questions posed in a multiple-choice format. There is less agreement, however, on why this is the case, and most attempts at explanation, with the possible exception of the work on guessing behavior by sex, appear to be little more than ad hoc speculations. Some of these speculative hypotheses are cognitively essentialist in form, proposing general differences between males and females in the ability to respond to questions posed in a multiple-choice format. Others are more social-psychologically grounded and include a role for the subject-matter being assessed. Still other proposed explanations are complex combinations of cognitive and social-psychological factors (e.g., the explications of strategies for guessing which differ by gender). Though these speculations serve to continue to document the existence of performance differences by sex, they offer little in the way of explanation. For all the research and speculation, to the extent that males and females perform differently on questions presented in a multiple-choice format, the question of why the differences occur is still unanswered.

The Assessment of Achievement in History

Although this study focuses on the relationship between epistemic conceptions and academic performance, the assessment of performance is not without its own context. In this case, it is clearly embedded within the discipline of history. I do not take this sense of contextualized assessment as calling for an extended survey and analysis of history as an academic discipline; however, it is worthwhile to consider other reports of assessments in history, keeping a watch for references to possible difference in performance by gender.

The US History Report Card, prepared by the Educational Testing Service as part of the National Assessment of Educational Progress (NAEP), reported the achievement of fourth, eighth, and twelfth-grade students in history (1990). The assessment measured knowledge and understanding of various historical periods, events, persons, and contexts. There were four levels of proficiency in the assessment, beginning with the ability to identify simple historical facts and moving through the interpretation of historical information and ideas. The assessment was composed primarily of multiple-choice questions (though twelfth-grade students were asked to write about specific historical periods and relationships). Although a large majority of high-school seniors (89 percent) showed a knowledge of historical information, only 5 percent performed at the highest proficiency level, which required students to demonstrate understanding of the interrelationships among historical events, persons, and documents.

For questions asked in a multiple-choice format, there was a difference by sex on the NAEP assessment with regard to (1) responses to specific questions, and (2) the overall level of achievement attained. For example, male students outperformed females in identifying the historical context of the first atomic bomb, while females scored higher on a question about Susan B. Anthony. An achievement level of 350 indicated knowledge of social, cultural, religious, and regional diversity, as well as an understanding of ideals and traditions such as equality, opportunity, and tolerance. Such cultural and social content (assuming it reflected an attempt to understand women's lived experiences and was not limited to androcentric treatment of topics such as work and family) could be

thought to be more interesting to female students. Yet males achieved proficiency at this level at more than twice the rate of females.

In a study of performance on the 1986 AP US History examination, Breland and his colleagues considered a number of factors which could have contributed to the higher performance of males on multiple-choice tests and the slight advantage shown by females on tests composed of essays and other constructed answers (Breland, Danos, Kahn, Kubota, and Sudlow, 1991). The researchers examined the results of a randomly selected sample of male and female students from the population who took the 1986 examination. Males significantly outperformed females on the multiple-choice part of the examination. Females performed at a level non-significantly above males on the constructed-response section (standard essay plus document-based question) taken as a whole.

The Breland study rated the constructed responses in the sample on the dimensions of composition, historical content (including factual errors), handwriting quality, neatness, and number of words written. The findings showed that composition quality and number of words written were the best predictors for scores on the document-based question. Ratings of historical content and scores on the multiple-choice part of the examination predicted best for scores on the traditional essay question. Handwriting and neatness showed a relatively low correlation with both types of constructed-response questions.

Breland proposed that the multiple choice items may assess different skill constructs than constructed-response item types. Multiple-choice questions appeared to assess a broad range of historical knowledge. The two constructed response questions types, while requiring composition skills and verbal fluency, appeared to assess a narrower

range of historical knowledge. The document-based question in particular, according to Breland, may require more English-related skills and relatively less historical knowledge than conventional essay questions. Breland suggested that males may have more knowledge of history, though such a conclusion would seem to rest on a conventional definition of knowledge as historical coverage.

Bridgeman and Morgan (1993), in a study comparing the achievement of two groups of high school students who took AP examinations in US and European History with scores on standardized achievement tests and achievement in college courses, analyzed AP performance for students from 38 colleges across ethnic groups. In every group, women's essay scores were higher than their multiple-choice scores. Men showed higher multiple-choice scores in every ethnic group with the exception of the small group of Puerto Rican males. Aggregated across ethnic groups, males and females performed at essentially the same level on the essays, while males outperformed females on the multiple-choice part of the test.

Within the domain of history, it is also important to consider the nature of the historical content being assessed. For AP US History, the content of the course (and examination) is relatively stable over time. At present, the development committee for the course has identified five content categories to guide the development of curriculum and teaching activity. The multiple-choice part of the AP test reflects the content emphasis in effect at any given time. The current categories and their associated weightings on the examination are: political institutions and behavior/public policy (35%), social change

(35%), diplomatic and international relations (15%), economic developments (10%), and cultural and intellectual developments (5%).

For at least the last ten years, males have outperformed females on the multiple-choice section of the AP US History examination by approximately one-third of a standardized difference score. For the same period, females have performed approximately the same as males on the document-based questions, at times doing slightly less well than males and in other years doing noticeably better.

It is worth noting that on the document-based questions, there is a difference in performance by gender in relation to the content category of the question. Females have earned higher scores on DBQs that have high social and economic content. Males have done well on questions with high political and diplomatic content. And though these findings cannot be claimed as evidence for a general causal effect of content on achievement -- there is an overlap of content across multiple-choice and the topic of the constructed-response document-based question, yet there is differential performance by item types across sex lines -- the nature of historical content clearly interacts with sex in the process of academic assessment.

Summary and Conclusion

A concluding statement is appropriate at this point. At the beginning of the chapter, I stated my intention to review the work of researchers who have conducted gender-related research that might offer alternative explanations to the hypotheses I have proposed. I also indicated that the literature review would help substantiate the need for my research. By way of providing a summary, I will revisit that claim of substantiation.

I would like to consider the last section of the literature review first. The recent work carried out at ETS on the relationship between the design of assessment and subsequent performance has clearly found performance differences by sex across item types. These findings reinforce my claims regarding the existence of such differences.

The ETS work also considered the nature of domain-specific knowledge in performance. Breland, et al. (1991) suggested that the reasons males perform better on multiple-choice questions on the AP US History examination is because they know more history. Females, according to Breland, do at least marginally better on at least some types of constructed response questions because they are better writers--better in the sense of having more developed skills in composition. I would argue, and this is at the heart of my study, that Breland's speculation is reflective or derivative of deeper epistemic differences.

It is my contention that a better explanation for the differences in achievement can be made by considering differences in how students think about the nature of historical truth-in-knowledge and how they organize information in general (historical or otherwise). In testing for mastery of particular historical content (e.g., knowledge of historical events and the structure of historical concepts), the type of assessment item may favor a student who thinks about truth-in-knowledge in one manner and organizes information using one cognitive style, and disadvantage a student who employs a different sense of truth-in-knowledge and organizes information differently. The overriding sense of the literature which has examined the nature of differences across sex and gender in learning provides sufficient grounding for developing an alternative approach to explaining gender-related

performance that goes beyond the simple descriptive (and circular) assertion that students of one sex know more history than students of the other.

Existing research has offered much evidence that all people, from early in their life, develop an idea of themselves as possessing a distinctive gendered identity. And while much behavior is no doubt similar across gender lines, much significant behavioral and cognitive activity is also undoubtedly different. With respect to schooling and education, there is much to distinguish between maleness and femaleness -- in the internalized motivations and aspirations children bring with them to school, in the expectations held for them by their teachers and their peers, in the preferences shown for different subjects. I would argue the case has been well made for the assertion that students participate in the process of formal education as gendered learners (Beal, 1994; Best, 1983). The work of Baxter-Magolda (1992) shows that for the majority of college students, academic learning is carried out in characteristic ways which differ between males and females.

But, the case for how students think about epistemological issues such as the nature of truth (truth-in-knowledge) and the organization of information, and what part such thinking plays in learning and performance, has not been made. There is a long-established strand of thought that holds women to be concerned with immediate context and men with longer-term goals; for women to prefer the sensate and men to favor abstract ideas. In her work articulating the moral development of women, Gilligan has drawn our attention to these differences. Belenky and her colleagues have extended these ideas directly into learning.

Yet there are differences between Gilligan and those who build on her ideas.

Gilligan sought to understand the cognitive activity that underlies decision-making in the moral domain. Belenky focused on approaches to learning. Gilligan operates in the realm of epistemological process and stance--how a normative belief is constructed and expressed. Belenky, et al., are instrumental--concentrating on how differences in approaches to learning might look in practice. Neither Gilligan nor Belenky addressed the question of which ideas about knowledge a learner brings with her into a classroom. The recent work of researchers such as King and Kitchner (1994) and Baxter-Magolda (1992) employs developmental categories similar to both Gilligan and Belenky, and extends their ideas by direct reference to a person's belief about the way the elements of knowledge come into existence at each level of cognitive development.

Bruner, on the other hand, talks directly about how a person thinks about the world in which they live. Though his conceptions of paradigmatic and narrative thought appear similar, at least descriptively, to the conceptions of separate and connected knowing, Bruner is not interested in questions of why or who--he is silent on how we come to prefer one mode over the other, or whether a person's gendered development contributes to their preferring to view the world as having primarily either a narrative or paradigmatic structure.

Considered from the classical perspective of epistemology, Bruner is operating at a place before justification and truth. And though he doesn't speak to the issue himself, it is not unreasonable to think that knowledge claims (or to use the language of my research, beliefs about truth-in-knowledge) can develop from either mode of initial cognitive

organization of information. The particular mode of organizing information, narrative or paradigmatic, can be thought of as providing the cognitive substrate from which subsequent knowledge claims develop. While the concept of cognitive orientation does not providing criteria for justification or truth, it is not unreasonable to think that a characteristic mode of cognitive orientation might be related to particular truth-in-knowledge positions (e.g., there may be a relation between a paradigmatic cognitive orientation and believing truth-in-knowledge to be absolute). The grounding for such possible relationships are examined in the research.

The discussion above leads to the following summary of the description and purpose for my research. The research will focus on: (1) documenting the existence of a gender-linked conception of the nature of historical truth (truth-in-knowledge), and (2) documenting the existence of a gender-linked narrative or paradigmatic cognitive orientation. I also will: (3) examine the relationship of these epistemological and cognitive conceptions with categories of sex and gender and the types items of assessment as they interact in one academic context.

The purpose of the research -- at least in the short run -- is to gain a better understanding of why the sex-related differences in performance on the AP US History examination have come to be and continue to exist as they do. The ultimate justification of such research will be its contribution to meeting the challenge of developing equitable conditions across the landscape of learning for all students regardless of the differences they bring to school and all the other places where learning occurs.

Chapter 3

The research design developed for the study utilized cluster sampling of schools and intact classes to create a sample of 190 students from 14 schools in two states. Random sampling of schools from one large and one small state insured that schools in the study were representative by size and location of schools currently offering courses in AP US History in the New England region. Students in classes selected into the sample completed four instruments: (a) the Nature of Historical Explanation and Truth questionnaire; (b) the Narrative/Paradigmatic Organization of Knowledge questionnaire; (c) the Bem Sex-Role Inventory; and (d) the Student Biographical questionnaire. The overall score and scores for the three sections of the 1997 AP US History examination served as outcome variables in the study. The purpose of the design was to test 19 hypotheses which focused on the possible existence of sex and gender-linked conceptions relating to historical truth and cognitive orientation, and to examine the extent to which such conceptions contribute to differences in performance by sex on the AP US History examination.

Sample

Students from three-quarters of the 1,074 public and non-public high schools in New England take one or more Advanced Placement examinations each year. During the May, 1997 examination period, students at 806 schools in the six-state region took the AP examination in US History. Of the schools that offer AP examinations in US History in New England, approximately three-quarters (74.1%) are public (state and local school-

board) institutions. Students from these public high schools accounted for 70.1% of the region's AP US History examinations in 1997.

Though the number of students attending religiously-affiliated or non-religious independent schools may be greater in New England than in other parts of the country, the number of such schools in the total population of secondary schools is still small in relation to the public schools. In 1997, 82.2 % of the AP US History examinations given nationally were taken by students enrolled in public high schools. Data from classes selected from among public high schools are, therefore, generalizable to the population of AP US History test-takers, and the schools selected for the study provided the most representative sites of instruction for this study.

The research sample consisted of students who, in the 1996-1997 school year, were enrolled in AP US History classes in two representative New England states, Massachusetts and New Hampshire. A complete list of individual AP US History classes (course sections) in both states was not available. Therefore, the unit of sampling chosen for the study was the school.

Though students may take AP examinations any time during their academic career, most examinations are taken by students completing the last year (grade 12) of secondary school. AP US History is the exception to this norm. Eighty-five percent (85.2 %) of the AP US History examination cohort in 1997 were enrolled in grade 11 at the time they took the test. This extensive testing in grade 11 is probably due to the existence of a required American History course in the junior year in most secondary school curricula. The majority of schools that offer an AP course in US History offer it as a one-year course

in place of an advanced or honors section of the grade 11 American History course. To support comparability across students, the sample selected for analysis included only AP test-takers who were completing grade 11 in the spring of 1997.

The sampling procedure for the study was designed to yield a sample of approximately 200 students. Because it was not possible to know the exact number of students in a class who would agree to participate in the study and would in turn take the AP examination, and because random selection could result in the inclusion of schools with unexpectedly small classes in the sample, the decision was made to select a sample of AP US History classes which, based on estimates of the percent of students who would agree to participate, past average class sizes, and test-taking rates, would yield a final sample of approximately 200 students.

To insure representativeness across geographical location, the sample of schools was drawn from the combined group of public schools in Massachusetts and New Hampshire which tested students in AP US History in 1996. Using an approximate average of 15 AP US History examinations per public school as a guide, and keeping in mind the need to oversample to allow for non-participation and non-test taking among students, the decision was made to choose fifteen schools for participation in the study. The proportion of schools to be selected from each state was then determined by comparing the number of AP US History examinations given in each in 1996. The number of schools chosen from Massachusetts was thirteen. Two schools from New Hampshire were selected.

If the building principal either declined to participate or was otherwise unable to be included in the study (e.g., AP US History was not offered in 1996/1997), contact was made with the next school on the list. Seven schools in Massachusetts were chosen for participation but the principal either declined, or was unable, to participate. The reasons given for non-participation were related to the time needed to comply with the requirements of the study. One school was included in the study, but no data were included in the analysis because only information from grade 12 students was provided. To insure representation across the largest possible number of schools, if a school offered more than one class (section) of AP US History, teachers were asked to select one class at random for inclusion in the sample.

At the classroom level, students who did not wish to participate were excused from the study. Three declined to participate, indicating their choice on the parent/student consent form provided to participating teachers. 40 students were excluded from the sample because they were in grade 12 during the 1996-1997 school year. An additional 53 students did not complete one or more of the four instruments administered. Information provided by these students was not included in the subsequent analyses. The final number of students in the sample was 190.

Instrumentation

Nature of Historical Explanation and Truth questionnaire. The Nature of Historical Explanation and Truth questionnaire (NHET) is a subject-specific, nine-item questionnaire developed to assess the extent to which students believe historical truth: (a) exists and is knowable with certainty, (b) does not exist because of the relative nature of

the process of determining historical explanation, or (c) exists conditionally as a socially-constructed construct subject to change (see Appendix A). The questionnaire consists of nine descriptive scenarios drawn from American history that are commonly encountered in an AP US History course. A question of historical explanation follows each scenario. Respondents were not asked to think about the answer to the question itself. Rather, they were instructed to select the statement that best described the sense of historical truth (a, b, or c, above) which would emerge from the process of attempting to answer the question.

The nine items (scenarios and questions) on the questionnaire were suggested by two experienced AP teachers. The scenarios were chosen from different historical periods and represented social as well as political historical content. The questions were reviewed and rewritten to eliminate possible normative tone -- that is, questions that might require explanations based on moral or religious values were not included in the final version of the questionnaire. The instrument was pilot tested in two AP US History classes. A review of the student responses and discussion with students indicated that students understood the instructions and intent of the task.

The following historical scenario and accompanying question provides an example of the items included in the NHET:

Andrew Johnson assumed the presidency under difficult conditions -- a terrible civil war had just come to an end and the country had suffered the assassination of a sitting president. Though born in the South, he had supported the Union cause during the war. His poor background left him suspicious of both well-to-do whites and newly-freed Negroes. He engaged in a running battle with Congress for much of his time in office. Many felt he would be the first president to be removed through the process of impeachment.

To what extent were President Andrew Johnson's troubles due to his own actions versus the political climate of the times?

The proportion of the total responses across the nine items that were entered for each of the three definitions of historical truth was calculated for each respondent.

Students who did not answer at least five questions on the questionnaire did not have their choices recorded and were excluded from the dataset.

Narrative/Paradigmatic Organization of Knowledge questionnaire. Using the conceptual framework proposed by Bruner, a 10 item open-ended questionnaire (N/POK) was developed to assess relative preference for using narrative and paradigmatic cognitive modes for organizing perceptions and experience. The questionnaire was designed to assess cognitive activity independent of any particular subject-matter domain, with the stem scenarios chosen to reflect non-academic events which would be familiar to students in grade 11 (see Appendix B).

The following question is drawn from the N/POK questionnaire:

You've been selected to be on a committee to make recommendations to the school board about whether uniforms should be required for all students in the schools of your district. Up to now, you haven't given much thought to this matter. But to your relief, another member of the committee gives you the numbers of several schools that had discussions about school uniforms last year. As you get ready to call these schools (using the school board's credit card!), what are five questions you would ask to help you prepare for the committee's first meeting?

Partial (five-item) versions of the questionnaire were pilot-tested in four high schools in Massachusetts in classes of advanced and AP US History students. A review of the responses collected from the four classes showed that students were able to understand the scenarios and construct meaningful responses as instructed. The analysis

of the pilot data revealed that one question was not providing the opportunity for students to respond with a mix of narrative or paradigmatic responses. The question was rewritten and included in the final form of the questionnaire, though the question was ultimately discarded from coding when inspection of a set of responses drawn from the study sample showed that students were having difficulty understanding the question and were either not completing it or were giving non-codable responses.

The final nine-question form of the questionnaire was subsequently tested for interrater reliability. A coding guide was prepared (see Appendix C) and was reviewed by an experienced coder. Following a discussion of the guide with the coder, a sample of 120 responses chosen from four questions administered in the pilot testing was coded and the results compared with the coding of the investigator. A set of 30 complete (nine-item) questionnaires drawn from the final sample was then coded by the investigator and the coder. The comparison of coding on the sample of thirty complete cases showed an agreement of 85.9% and a Cohen's Kappa statistic of .72.

The questionnaire was scored by calculating the percent of responses which were coded as paradigmatic and the percent coded as narrative across the responses given to all nine questions. Students who did not respond to at least five questions were not coded, and were excluded from the analyses of the sample dataset.

Bem Sex-Role Inventory. The Bem Sex-Role Inventory (BSRI) was designed to assess the extent to which individuals select as self-descriptive a range of socially desirable personality traits and behaviors that have been culturally defined as either masculine or feminine. The short form of the BSRI (the version used in the study)

contains 10 personality characteristics judged to be more desirable for females than males, 10 that were judged more desirable for males, and 10 gender-neutral descriptions. The short version of the BSRI was chosen because the items included were more likely to correspond to socially desirable gender-linked characteristics in a contemporary social context (e.g., the short form of the BSRI deletes several femininity items that were relatively socially undesirable, such as “gullible” and “flatterable”).

Standard-setting for the original version of the inventory was carried out in 1972 and 1973, using 100 undergraduate students at a selective university as judges of a list of 400 personality descriptors, 200 of which were deemed positively valued and stereotypically feminine or masculine. Traits that were judged by both female and male judges to be significantly more desirable for each sex were included in the final pool of characteristics, with 20 items ultimately selected for the masculine and feminine scales. The two scales were similar in the mean social desirability rating of the 20 items. Twenty other descriptive characteristics which were judged no more desirable for either sex were selected as filler items for the inventory.

Individuals completing the BSRI-short form are asked to indicate on a 7-point scale how well each of the 30 characteristics describes himself or herself. Using the median-split method of scoring the inventory (in which a respondent is scored as above or below the median of a reference group for each scale) it is possible to score high on both female and male scales (an androgynous condition), to score low on both (undifferentiated), or high on one dimension and low on the other (either feminine or masculine).

Depending on the median-split placement on each sex-role scale, a respondent is classified as either self-described feminine (F), masculine (M), androgynous (A), or undifferentiated (U). In Bem's 1975 sample of 375 males and 290 females at Stanford University, the median-split procedure resulted in the following percentages of male and female subjects classified as feminine, masculine, androgynous, and undifferentiated, respectively: males -- 16%, 37%, 21%, 27%; females -- 34%, 16%, 29%, and 20%.

While the BSRI has come under criticism in the years since its development and introduction in the late 1970's (see Spence & Helmreich, 1981) it continues to be widely used, and the chief objection of critics like Spence -- that Bem is conceptually inconsistent in her use of the BSRI in the development of gender schema -- is not relevant when there is explicit recognition of both masculine and feminine dimensions at work in the creation of a person's sex-role orientation (use of the median-split method of scoring the BSRI assures this). The BSRI received a favorable review in *Measures of Personality and Social-Psychological Attitudes* (Lenney, cited in Robinson, Shaver, & Wrightsman, Eds., 1991), with the only caution being against extending use beyond its original theoretical intent. In the current study, placement on the BSRI was used as a comparison with sex categories when investigating differences in academic performance, a use that does not violate the caution noted above. See Appendix D for a copy of the instructions for administration and five descriptive personality characteristics selected from the short form of the BSRI.

Student Biographical Questionnaire. Students were also asked to complete a 7-item survey which was designed to collect attitudinal (dispositional), ability, and

achievement data (see Appendix E). The SBQ included three questions oriented toward assessing attitude toward history as a discipline, asking students: (a) how they felt about history compared with other subjects, (b) if they planned to take history courses in the future, and (c) how they felt about the role of history in the development of their world view. Responses to these questions were selected from among three ordinal-level choices (low to high). Developed ability was measured by collecting self-reported PSAT verbal and mathematics scores. Achievement was assessed by asking students to indicate their GPA, to estimate the grade they would most likely receive in the AP US History course, and to judge how well prepared they felt before taking the AP US History examination.

The Advanced Placement examination in US History. The score received on the AP US History examination and the sub-scores achieved on the three sections of the examination were selected as dependent (outcome) variables for the study. Though the Advanced Placement examination in US History has changed in form across the years, the changes made by the test development committee are deliberately infrequent, usually reflecting a shift in the weighting of one part of the examination or a reallocation of time given to different sections of the test. The 1997 examination was three-hours and five minutes in length. It consisted of two sections: (a) a 55 minute multiple-choice section, and (b) a 130 minute free-response section. The historical content assessed by both multiple-choice and free-response questions is weighted as follows: political institutions and behavior and public policy, 35% of the examination; social change, 35%; diplomacy and international relations, 15%; economic developments, 10%; and cultural and intellectual developments, 5%.

There are 80 questions on the 55 minute, Section I multiple-choice (M-C) part of the examination. Questions on the multiple-choice section are designed to test factual knowledge, breadth of preparation, and knowledge-based analytical skills.

Section II of the examination begins with a mandatory 15 minute reading period, followed by Part A, where students are required to answer a document-based essay question (DBQ) in 45 minutes; and Part B, in which students are given 70 minutes to answer two standard-format essay questions (ESSAY) chosen from two sets of two questions each. The essays require students to relate developments in different areas, to analyze common themes in different time periods, and to compare individual or group experiences that reflect socioeconomic, ethnic, racial, or gender differences. Answers to the essay questions are judged on the strength of the thesis developed, the quality of the historical argument, and the evidence offered in support of the thesis, rather than simply on factual information.

The DBQ differs from the standard essays in its emphasis on the ability to analyze and synthesize historical data and assess verbal, quantitative, or pictorial materials as historical evidence. Teachers and students are notified a year in advance of the 50-year period from which the DBQ will be constructed. See Appendix F for examples of standard-essay and document-based essay formats for the AP US History examination.

The total score on the examination (TOTAL) is determined by adding the scores on the multiple-choice section (adjusted for guessing), the two standard essays, and the document-based questions, with the MC score counting for one-half of the total score. Within the free-response part of the examination (Section II), the standard essays (taken

together) count for 55%, the DBQ 45%. A final score of 1 to 5 is then awarded by the chief faculty consultant for the examination. A score of 1 is considered “least qualified” (not recommended for credit); a 5 is considered “extremely well qualified” for credit or placement. Most colleges give advanced placement or credit for scores of 3 or above. In 1997, 54.7% of all test-takers received scores of 3 or above on the AP US History examination. The mean score earned was 2.84 ($SD = 1.20$).

Procedure

In September, 1996, principals of the schools selected for the study were contacted by telephone to gain permission to talk with the AP US History teacher regarding participation in the study. Once the approval of the principal was received, the teacher was contacted by telephone. After both the principal and the teacher indicated their willingness to participate, they were sent letters which briefly described the study and asked to sign and return a copy of the letter indicating their final agreement to become part of the study. Both principals and teachers knew that the investigator was employed by the College Board (sponsor of the AP program), but requests for participation were made on the basis of the study being a part of dissertation research being conducted independently of other College Board or ETS research.

The teachers were then asked to discuss the study with their students and to solicit participation from them. They were asked to report back if less than 90% of the class agreed to participate. None did so -- though in fact the ultimate student participation rate was estimated to be less than 90% in two of the schools. Students and their parents (for students less than 18 years old) were asked to sign a consent form to indicate their

agreement to participate in the study. The consent forms were collected during the period from November, 1996 to June, 1997. Incomplete consent forms were returned to the teachers for completion by students or parents.

In the spring of 1997, after the final sample of schools and classes was determined, a letter was sent to the participating teachers with information regarding the administration of the four instruments which made up the research battery. Following the letter, calls were made to set up appointments to deliver sets of the research battery to the schools. All materials were hand-delivered to the 15 participating schools after the administration of the AP US History examination (May 9th) during a ten day period in mid-May, 1997.

Teachers were asked to have students complete the four instruments -- the NHET, N/POK, BSRI, and SBQ -- as much as possible during instructional time. Informal discussion with participating teachers indicated that approximately 80% of the students completed the questionnaires and surveys in class. The time needed for each student to respond to the questions on all four instruments (including orientation and time for questions, as well as actual writing time) was approximately 60 minutes. The collection of data was completed by late June, 1997. Data transformation, coding, and the development of a dataset was completed in August, 1997.

As noted, students in the sample took the AP US History examination in May, 1997. As part of the agreement to participate in the study, students gave permission for their overall and multiple-choice and free-response section scores to be released for research purposes. These data were forwarded from ETS in July, 1997 and incorporated

into the dataset through matching with the names and ETS identification numbers of students from who had completed the four research instruments.

The dataset included a record for each student in the sample who took the AP US History examination and completed all four of the instruments in the research battery (the decision rule used for inclusion was that the student receive a code for the BSRI, and respond to the majority of questions on the other three instruments). In addition to identification information (e.g., student name and AP identification number, school identification number, year in school) each student record contained data from the following individual and questionnaire-based variables:

1. Student sex (Sex). Either male or female self-designations were used.
2. Bem Sex Role Inventory designation (BSRI). According to responses given on the BSRI, each student received one of the following four designations of self-described gender preference: feminine (F), masculine (M), androgynous (A), or undifferentiated (U).
3. Nature of Historical Explanation and Truth (NHET). The values recorded were the percent of responses given to each of the three variables which described the nature of historical truth.
4. Narrative/Paradigmatic Organization of Knowledge (N/POK). The values recorded were the percent of responses coded as narrative and the percent coded as paradigmatic.
5. Score on the AP US History examination (TOTAL). The total score on the 1997 AP US History examination achieved by each student.
6. Score for the multiple-choice section of the examination (M-C).
7. Score for the document-based question section of the examination (DBQ).

8. Score for the standard essay part of the examination, summed over the two questions chosen (ESSAY).
9. Grade-point average (GPA). Self-reported across all courses.
10. Estimated course grade (GRADE). Estimated grade to be received in the AP US History course.
11. Estimate of quality of preparation for the AP US History examination (EXAMPREP).
12. PSAT verbal score (PSAT/V). Self-reported score for each student.
13. PSAT quantitative score (PSAT/M). Self-reported score for each student.
14. Favorability toward history in comparison to other high school subjects (LIKEHIST).
15. Intention to take other history courses in college (TAKEHIST).
16. Importance of history in the development of a world view (HISTIMPT).

Data Analysis

The predictions that student sex and sex-role orientation are related to: (1) beliefs regarding the nature of historical explanation and truth, (2) preferred mode of cognitive organization, and (3) performance on different sections of the AP US History examination, was operationalized by first carrying out an analysis of the descriptive data gathered for students in the sample. Following this, two sets of zero-order correlational analyses and a group of 12 multiple-regression models were created. Lastly, 12 additional models using R^2 -change values generated from multiple regression and canonical correlation analysis to estimate explained variance between and among predictor and outcome variables were also created.

Descriptive statistics. A set of descriptive statistics was generated for the following variables: (a) BSRI, (b) GPA, (c) anticipated grade in AP US History (GRADE), (d) quality of preparation for the AP examination (EXAMPREP), (e) favorability of history compared with other high school subjects (LIKEHIST), (f) possibility of taking additional history courses in college (TAKEHIST), (g) importance of history in the development of a world-view (HISTIMPT), (h) PSAT/V, (i) PSAT/M, (j) responses to the nature of historical explanation and truth questionnaire (NHET), (k) narrative/paradigmatic orientation (N/POK), and (l) total and section-specific performance on the AP examination in US History. For interval-level variables, mean and SD data were generated for female and male subsets of the total sample. Tests for differences between males and females (t-tests) were made for interval-level variables. Tests of association for categorical-level variables were made using chi-square as a statistical measure.

For the analysis of descriptive data, it was hypothesized that:

1. There would be no difference between females and males in the distribution of the values reported for the following achievement variables: (a) GPA, (b) anticipated AP US History course grade, and (c) perceived quality of preparation for the AP examination.
2. There would be no difference between females and males in the distribution of the values reported for the following dispositional variables: (a) favorability of history compared with other high school subjects, (b) importance of history in the formation of a world-view, and (c) likelihood of taking additional history courses in college.

3. In responses to the NHET questionnaire, females would show a higher percentage of historical truth as relative (Truth/R) responses than males.
4. In responses to the NHET questionnaire, males would show a higher value on the historical truth as absolute (Truth/A) variable than females.
5. In responses to the N/POK questionnaire, males would have higher percentages of responses coded as paradigmatic than females.

Intercorrelation of data. The second statistical analysis consisted of two sets of correlation tables. The first set was three correlation matrixes which showed the intercorrelations among: (a) BSRI, (b) Sex, (c) PSAT/V, (d) PSAT/M, (e) responses to the NHET questionnaire, and (f) responses to the N/POK questionnaire. Correlation matrixes were generated for the total sample, for female students, and for male students.

A second correlational matrix examined the relationship among the variables listed above with the four measures of performance on the AP US History examination. The measures of performance associated with the AP examination were: (a) total examination score (TOTAL), (b) score on the multiple-choice section (M-C), (c) score on the document-based essay section (DBQ), and (d) score on the standard-essay (ESSAY) section.

These two correlations permitted the assessment of the zero-order correlational relationships among the three measures of cognitive and sex-role orientation (orientation toward the nature of historical truth, the organization of knowledge, and the BSRI) and performance on the AP US History examination as a whole and its three constituent parts.

For the analysis of zero-order correlational data it was hypothesized that:

6. There would be a significant correlation between historical truth as relative (Truth/R) and DBQ scores.

7. There would be a significant correlation between historical truth as absolute (Truth/A) and M-C scores.

8. There would be a significant correlation between paradigmatic cognitive orientation (P) and M-C scores.

9. There would be a significant correlation between feminine sex-role orientation on the BSRI (F) and DBQ scores.

10. There would be a significant correlation between masculine sex-role orientation on the BSRI (Bem M) and M-C scores.

Multiple regression models. Multiple regression models were constructed for each of the four dependent variables -- the total score and the three scores for the multiple-choice, document-based question, and standard essay sections of the examination. The variables listed in the section on correlational analysis were used as independent variables. Three regression models were created for each of the four dependent variables - one for the total sample, one for female students, and one for male students. To aid in the interpretative value of the models, the school a student attended was included in the multiple regression analyses as a dummy-formatted block variable. The overall explanatory effect of the model (R^2) therefore included the contribution of the School variable.

Examination of the beta-weight values obtained from the multiple regression models contributed to understanding, for the total sample and for each sex, the relative

importance of each independent variable in predicting scores for the AP US History examination as a whole and for scores on each of the three item types that constitute the examination. Summary R^2 values for each of the models was also determined. In multiple-regression analysis, R^2 values represent the percentage of variance in the outcome variable accounted for by the set of independent variables (including, in this case, the effect of the School variable entered as a block in each model. See above).

For the multiple regression models, it was hypothesized that:

11. Truth/A would be a significant predictor of M-C scores for males.
12. Paradigmatic cognitive orientation (P) would be a significant predictor of MC scores for males.
13. PSAT/M would be a significant predictor of MC scores for males.
14. Truth/R would be a significant predictor of DBQ scores for females.
15. Narrative cognitive orientation (N) would be a significant predictor of DBQ scores for females.
16. PSAT/V scores would be significant predictors for performance on the standard essay section (ESSAY) of the examination for (a) males and (b) females.

R^2 -change models.

Using the same set of predictor and outcome variables used in the multiple regression models, a set of four models (one for each outcome variable -- each model contained R^2 values for the total sample, female, and male students) was created to show the amount of change in variance in intermediate and outcome variables accounted for by individual and sets of other intermediate and exogenous variables (The term exogenous is

used as a modifier for an actual or conceptual variable which is assumed not to be influenced by other variables within the model. Intermediate variables are assumed to be influenced by variables within the model.) Though the R^2 -change models developed for the study are not analyzed according to path analysis methods, the terms describing the variables is a reasonable approximation of how such terms are used in path analysis and are therefore used here. The indication of variance accounted for was the R^2 -change value, the additional variance gained when a variable is entered last in a list of regression predictors or when the squared canonical coefficient between two sets of variables is determined.

For the R^2 -Change models, it was hypothesized that:

17. Bem category (BSRI) would account for a significant portion of the variance in cognitive orientation for (a) females and (b) males. This condition was predicted to hold across all outcome variables.

18. Cognitive orientation would account for a significant portion of the variance in nature of historical truth for (a) females and (b) males. This condition was predicted to hold across all outcome variables.

19. Conceptions about the nature of historical truth would account for a significant portion of the variance for all outcome variables for (a) females and (b) males. This condition was predicted to hold across all outcome variables.

Chapter 4

Descriptive Statistics

Performance on the AP US History examination. T-tests for differences across sex were carried out for the overall score on the AP US History examination and for the three examination sub-scores. For the overall test score, there was a significant difference (favoring males) in the scores of females and males. The multiple-choice (MC) section of the examination also showed a significant difference across sex, again favoring males. Scores for females on the document-based question (DBQ), though higher, were not significantly different from those of males. The standard-essay scores (ESSAY) were significantly different in favor of males (see Table 1).

Sex and Bem category. Students in the study were assigned a Bem category designation depending on their responses to the Bem Sex-Role Inventory (BSRI). While both males and females were coded into all four Bem categories (e.g., students of both sexes were classified as Bem-Feminine, Bem-Masculine, Bem-Androgynous, and Bem-Undifferentiated), the distribution across Bem category was not the same for males as it was for females. Considering the sexes separately, females were more likely to be coded androgynous and masculine, and less undifferentiated and feminine, than females in the original Bem norming sample. The overall difference between the study sample and norming sample was significant². Male students were not significantly different from the original Bem sample in their Bem-category distribution (see Table 2).

² An alpha level of .05 was used throughout the study as a significance level and was highlighted with an asterisk in the tables and figures.

Table 1

Comparison of Female and Male Distribution Across Interval-Level
Independent and Outcome Variables

Variables	Females		Males		t	df	p
	M	SD	M	SD			
AP US History Examination							
Total Score	3.09	1.25	3.46	1.19	-2.07	188	.040*
Multiple Choice Score	38.29	16.13	44.49	13.89	-2.80	188	.006*
DBQ Score	4.94	1.68	4.50	1.56	1.86	188	.064
Essay Score	7.83	2.91	8.68	2.82	-2.02	187	.044*
Anticipated Course Grade	3.34	0.58	3.38	0.50	-0.54	179	.589
PSAT Verbal	62.23	9.56	61.67	7.59	0.44	188	.662
PSAT Mathematics	60.35	9.86	64.15	8.02	-2.87	188	.005*
Historical Truth							
Relative	.30	0.20	.32	0.15	-0.66	188	.513
Constructed	.45	0.21	.43	0.16	0.68	188	.500
Absolute	.25	0.15	.25	0.15	-0.06	188	.954
Cognitive Orientation (Para)	.60	0.01	.59	0.11	0.77	188	.443

Note. Negative t-value favors males. *p<.05.

Table 2

Sex and Bem-Category of Subjects

Bem Category	<u>N/n</u>	<u>P</u> (Study Sample)	<u>P</u> (Bem Norming Sample)
Female Students	106		
Feminine	23	21.7	23.8
Masculine	25	23.6	15.6
Androgynous	43	40.6	37.1
Undifferentiated	15	14.2	23.5
Male Students	84		
Feminine	6	7.1	16.0
Masculine	34	40.5	32.6
Androgynous	17	20.2	23.9
Undifferentiated	27	32.1	27.5

Note. For female students across Bem categories compared with Bem-female norming sample, $\chi^2(3, N=106) = 8.81, p = .032$. For male students compared with Bem-male norming sample, $\chi^2(3, N=84) = 6.85, p = .08$.

Table 3

Comparison of Female and Male Distributions Across Categorical-Level Descriptive Variables

Descriptive Variables	χ^2	<u>df</u>	<u>p</u>	<u>n/fem</u>	<u>n/male</u>
EXAMPREP	6.62	2	.037*	106	81
LIKEHIST	6.36	2	.042*	104	84
GPA	1.67	3	.645	104	84
HISTIMPT	1.28	2	.527	105	84
TAKEHIST	1.82	2	.403	105	84

Note. * $p < .05$.

Academic achievement. Students in the study showed a pattern of high academic achievement, as indicated by grades for prior coursework and scores on standardized tests. They also reported confidence in the likelihood of receiving a high grade in the AP US History course and satisfaction in the preparation they had received for taking the AP US History examination.

The first indicator of achievement was self-reported grade-point average (GPA). Students were asked to choose the category (from among four) that included their GPA. The first two categories were collapsed (the lowest category contained only one response) and a chi-square analysis was carried out to test for differences by sex across categories. There was no significant overall difference between males and females in the distribution of estimated GPA (see Table 3).

The second indicator of academic achievement was performance on the two sections of the Preliminary Scholastic Achievement Test (PSAT). Students were asked to report their scores for both the quantitative (PSAT/M) and verbal (PSAT/V) sections of the test. A t-test of PSAT/V scores showed no significant difference between females and males. A t-test of PSAT/M scores was significant, favoring males (see Table 1).

Students were also asked to indicate the grade they anticipated receiving in the AP US History course on a numerical scale with 4.0 as the highest grade. The mean anticipated grade for all students was 3.36. The difference in anticipated grade across females and males was non-significant (see Table 1).

A fourth measure of academic achievement was self-reported quality of preparation for the AP US History examination. Students were given a choice of three

categories from which to choose. The lowest category (“not very well” prepared for the examination) contained a small number of responses and was combined with the second category for the analysis. Males indicated receiving better quality preparation for the AP US History examination than females (67.9% vs. 54.7% in the “very well” prepared category). The chi-square analysis for differences in categories across sex reached a significant level (see Table 3).

The first two parts of the first hypothesis -- that there would be no difference between females and males with respect to (1) GPA, and (2) anticipated grade in the AP US History course -- were supported. The third part of the first hypothesis, the prediction that there would be no difference between females and males with respect to quality of preparation for the AP US History examination, was not supported.

Disposition toward History. Females in the sample were somewhat less favorably disposed toward History than males (15.4% of females, compared with 4.8% of males, indicated they “really don’t care for it” as a subject). There was a significant difference across sex in this indicator of disposition toward History (see Table 3), though overall the students in the study were favorably disposed toward History, with nearly 85% of the females and 95% of the males indicating they liked it “well enough” or “really liked it” compared with other subjects.

Students were also asked to indicate the importance of History in the development of their own world-view. For this indication of favorability toward History, there was no significant difference between males and females. Nearly 60% of both females and males indicated that History was “very important” to them in the development of their view of

the world. The χ^2 value for the distribution of responses to this variable was non-significant (see Table 3).

A third indicator of disposition toward History was the response given to the question regarding the possibility of taking a History course in college. Overall, slightly more than four or five subjects (82.5%) gave responses in the two highest categories, indicating they would either take more History courses in college or possibly major in History. The χ^2 test for differences between male and female responses was not significant (see Table 3).

The first part of the second hypothesis -- that there would be no difference between females and males regarding favorability of history compared to other high school subjects -- was not supported. The second and third parts of the second hypothesis, which predicted no differences between females and males with respect to the importance of history in forming a world view and the intent to take additional courses in history, were supported (see Table 3).

Nature of Historical Truth. For responses on the Nature of Historical Explanation and Truth instrument, females and males were virtually identical in their pattern of responses, with no significant difference found for t-tests between female and male responses for each of the three variables included in the instrument. (see Table 1). For the total sample (combining male and female responses), truth-as-constructed had the largest value (43.8%) among the three variables included in the NHET. Hypotheses 3 and 4, that women would consider the nature of historical truth as relative more than males,

and that men would consider historical truth as absolute more than females, were not supported.

Cognitive Orientation. For the total sample, slightly fewer than 60% of all responses on the Narrative/Paradigmatic Organization of Knowledge questionnaire were coded as Paradigmatic. Females and males were virtually identical in the proportion of Paradigmatic and Narrative responses (see Table 1). Hypothesis 5, that males would show a higher paradigmatic cognitive orientation than females, was not supported.

Intercorrelation of Data

For the correlation analyses, student sex was coded as 1 for female students and 0 for males. The BSRI variable was recoded as four dummy variables, with Bem F, Bem M, Bem A, and Bem U each being assigned the value of 1 if the student was assigned to the Bem category in question and 0 if they were not (e.g., a student coded as Bem U would have a value of 1 entered for the Bem U variable and 0 entered for the Bem F and Bem M variables). Both PSAT scores, verbal (PSAT/V) and mathematics (PSAT/M) were included in the set of predictor variables. The percent of responses coded as truth as relative (Truth/R), truth as absolute (Truth/A), and truth as constructed (Truth/C) were included in the set of predictor variables. Percent of responses entered as paradigmatic was included as the indication of cognitive orientation (Cog. Orien.).

Predictor variables. Zero-order intercorrelations among predictor variables for the total sample, and for female and male students, showed relatively few correlations between variables. For the total sample, aside from the artifactual correlations found among the dummy variables representing gender (e.g., the four Bem categories), among

sex and Bem categories, and within the nature of historical truth variables, the only significant correlations involved PSAT values. Scores for the two PSAT variables (PSAT/V and PSAT/M) were significantly correlated. PSAT/M was also significantly correlated with Bem A and with student sex. PSAT/V was significantly correlated with Truth/C (see Table 4).

The lack of intercorrelation among variables for the total sample was also seen for female students. In addition to the artifactual correlations noted above, there were only two significant correlations for females. PSAT/V showed a significant correlation with Bem U, and was also significantly correlated with Truth/C (see Table 5).

Males showed a similar pattern of intercorrelations. They displayed the same artifactual correlations as females. In addition, for males, the PSAT/M variable was significantly correlated with Bem A and Bem U (see Table 5).

Predictor and outcome variables. For the total sample, overall examination performance was significantly correlated with both PSAT/V and PSAT/M scores. Total examination score also showed a significant correlation with Bem A, Cognitive Orientation, and Sex. Multiple-choice scores were significantly correlated with PSAT/V and PSAT/M scores. MC performance was also significantly correlated with Bem A, Truth/A, Truth/C, and Sex. Correlation of document-based question scores with other variables reached significant levels with both PSAT scores and Cognitive Orientation. Essay scores were significantly correlated with both PSAT scores, Bem M, Bem A, Cognitive Orientation, and Sex (see Table 6).

For female students, the total AP examination score showed a significant

Table 4

Correlations Among Predictor Variables (All Subjects)

Predictor Variables	Bem F	Bem M	Bem U	Bem A	Sex	PSAT/V	PSAT/M	Truth/R	Truth/A	Truth/C	Cog. Orien.
Bem F	1.00										
Bem M	-.28*	1.00									
Bem U	-.23*	-.36*	1.00								
Bem A	-.29*	-.46*	-.36*	1.00							
Sex	-.20*	-.18*	-.22*	-.22*	1.00						
PSAT/V	.02	.03	-.05	.00	.03	1.00					
PSAT/M	-.05	.10	.12	-.18*	-.20*	.59*	1.00				
Truth/R	-.08	-.01	-.02	.10	-.05	-.09	-.01	1.00			
Truth/A	.00	-.01	.03	-.02	.00	-.13	-.07	-.36*	1.00		
Truth/C	.08	.02	-.01	-.08	.05	.18*	.07	-.68*	-.44*	1.00	
Cog. Orien.	-.08	-.04	.03	.07	.06	-.13	-.13	.00	.09	-.07	1.00

Note. *p<.05.

Table 5

Correlations Among Predictor Variables (Males and Females)

Predictor Variables	Bem F	Bem M	Bem U	Bem APSAT/	PSAT/M	Truth/R	Truth/A	Truth/C	Cog. Orient.	
Bem F	1.00	-.23*	-.19	-.14	-.03	-.19	-.05	.09	-.03	
Bem M	-.29*	1.00	-.57*	-.42*	-.08	.04	.04	-.03	.00	
Bem U	-.21*	-.23*	1.00	-.35*	.18	.30*	-.01	-.06	.06	
Bem A	-.43*	-.46*	-.34*	1.00	-.09	-.28*	.00	.05	-.05	
PSAT/V	.03	.12	-.24*	.04	1.00	.54*	.11	-.07	-.03	
PSAT/M	.07	.10	-.10	-.06	.64	1.00	-.02	.00	.01	
Truth/R	-.08	-.06	-.05	.16	-.17	-.03	1.00	-.42*	-.53*	
Truth/A	-.04	.01	.13	-.07	-.16	-.13	-.33*	1.00	-.54*	
Truth/C	.11	.05	.04	-.11	.28*	.12	-.74*	-.39*	1.00	
Cog. Orient.	-.14	.03	.14	.04	-.17	-.13	.01	-.01	-.01	1.00

Note. Values for males above diagonal; females below diagonal. *p<.05.

Table 6

Correlations Among Predictor and Outcome Variables

Outcome Variables	Bem F	Bem M	Bem U	Bem A	PSAT/V	PSAT/M	Truth/	Truth/	Truth/C	Cog. Orient.	Sex
All Students											
Total Score	.03	.09	.09	-.20*	.57*	.58*	-.07	-.09	.13	-.18*	-.15*
Multiple Choice Score	-.03	.09	.13	-.18*	.65*	.62*	-.09	-.15*	-.20*	-.08	-.20*
DBQ Score	.12	-.05	-.01	.03	.30*	.29*	.06	.01	-.07	-.22*	.13
Essay Score	.02	.18*	-.05	-.15*	.40*	.44*	-.04	-.02	.05	-.18*	-.15*
Females											
Total Score	.11	.08	-.03	-.14	.71*	.67*	-.05	-.07	.10	-.26*	
Multiple Choice Score	.06	.13	-.05	-.13	.73*	.67*	-.11	-.12	.19*	-.12	
DBQ Score	.13	-.05	-.03	-.04	.48*	.51*	.14	-.05	-.10	-.27*	
Essay Score	.05	.14	-.13	-.07	.59*	.54*	-.03	.01	.03	-.25*	
Males											
Total Score	-.03	.05	.16	-.22*	.38*	.40*	-.11	-.11	.21	-.08	
Multiple Choice Score	-.10	-.04	.25*	-.19	.54*	.49*	-.08	-.21	.26*	.00	
DBQ Score	.03	.00	.07	-.10	-.01	.01	-.06	.09	-.03	-.18	
Essay Score	.05	.18	-.05	-.20	.13	.24*	-.08	-.06	.12	-.08	

Note. *p<.05.

correlation with PSAT/V and PSAT/M scores and with Cognitive Orientation. Female multiple-choice scores were significantly correlated with both PSAT values and with Truth/C. DBQ and Essay scores showed significant correlations with PSAT/V and PSAT/M scores and with Cognitive Orientation.

Male students were similar to females in showing significant correlations between total examination scores and both PSAT/V and PSAT/M scores. Male total examination scores were also significantly correlated with Bem A values. MC scores showed significant correlations with both PSAT values, and were also correlated with Bem U and Truth/C. Male DBQ scores did not show significant correlations with any predictor variables. Male essay scores were significantly correlated with PSAT/M scores (see Table 6).

Hypothesis 6 predicted a significant correlation between historical truth as relative (Truth/R) and DBQ scores. Hypothesis 8 predicted a similar correlation between paradigmatic cognitive orientation and MC scores. Neither hypothesis was supported. Hypothesis 7 -- that there would be a significant correlation between truth as absolute (Truth/A) and multiple-choice score -- was supported (but opposite in direction to what was expected). Neither hypothesis relating Bem category for females and males to examination performance (hypotheses 9 and 10) were supported.

Multiple Regression Models

Multiple regression models using total score (TOTAL), multiple-choice questions (MC), the document-based question (DBQ), and essay questions (ESSAY) as dependent

variables were generated for the total sample, and female and male subsets. A total of 12 analyses was generated (see Table 7).

Total score. The analysis showed significant beta values for four predictor variables in the total sample: Bem M and Bem U, and PSAT/V and PSAT/M. The same four predictors had significant beta values for the subset of female students. There were no significant beta values found for predictor variables for male students. R^2 values for the three total score models using the total sample, female, and male subgroups showed high effect sizes and were statistically significant (see Table 7).

Multiple-choice. For the total sample, there were five predictor variables with significant beta values for the multiple-choice dependent variable. Bem M and Bem U beta values were significant, as were student sex (Sex), PSAT/V, and PSAT/M.

Females showed a pattern of significant beta values similar to the total sample. Excluding student sex (which was not entered in the multiple regression for female and male subgroups) the group of variables showing significant beta values was the same as for the total sample: Bem M and Bem U, and PSAT/V and PSAT/M. For male students, PSAT/V and Truth/A showed significant beta values (see Table 7).

Hypotheses 11, 12, and 13, which predicted that truth as absolute (Truth/A), paradigmatic cognitive orientation (P), and PSAT/M would be significant predictors of multiple-choice scores for males, were not supported.

For the three models which used total score as the outcome variable (total sample, females, males), R^2 values showed large effect sizes and were statistically significant.

Table 7
 Multiple-Regression Partial Correlation and Beta Coefficient Values for Predictor Variables
 Across Four Outcome Variables.

Outcome Variables	Sex		Bsm.F		Bsm.M		Bsm.U		PSAT/V		PSAT/M		Truth/B		Truth/A		Cog. Orient.		R ²
	Partial	B	Partial	B	Partial	B	Partial	B	Partial	B	Partial	B	Partial	B	Partial	B	Partial	B	
Total Score																			
All Subjects	-0.09	-.07	.14	.10	.20	.16*	.17	.14*	.24	.23*	.16	.15*	-.04	-.03	-.02	-.01	-.09	-.06	.61*
Females			.15	.09	.18	.11	.27	.17*	.36	.32*	.30	.25*	-.01	-.01	.04	.02	-.20	-.12	.76*
Males			.03	.03	.14	.16	.11	.13	.13	.14	.11	.12	-.17	-.16	-.14	-.13	.02	.02	.49*
Multiple-Choice																			
All Subjects	-.19	-.13*	.13	.09	.19	.14*	.22	.16*	.45	.41*	.16	.14*	-.09	-.06	-.09	-.06	.06	.03	.68*
Females			.13	.08	.24	.14*	.27	.17*	.45	.42*	.26	.22*	-.09	-.05	.01	.01	.03	.02	.75*
Males			-.04	-.03	.05	.05	.07	.07	.39	.38*	.16	.14	-.23	-.18	-.25	-.21*	.11	.08	.63*
Document-Based																			
All Subjects	.23	.22*	.07	.07	.02	.02	.06	.06	-.03	-.03	.07	.08	.13	.12	.10	.09	-.17	-.15*	.38*
Females			.10	.09	-.03	-.02	.11	.09	.09	.11	.21	.24*	.24	.21*	.08	.06	-.17	-.13	.54*
Males			-.06	-.07	.04	.05	.08	.11	-.12	-.14	-.09	-.11	.00	-.01	.15	.16	-.18	-.17	.32*
Essays																			
All Subjects	-.06	-.05	.08	.08	.20	.20*	.03	.03	.10	.11	.14	.15	-.03	-.03	.02	.02	-.07	-.06	.41*
Females			.01	.01	.15	.12	.10	.08	.26	.30*	.20	.21	.01	.01	.05	.04	-.14	-.11	.56*
Males			.14	.15	.25	.30*	.12	.14	-.09	-.10	.12	.13	-.11	-.11	-.10	-.10	-.00	-.00	.39*

Note. *p<.05. R² values are for overall models and include effect of Schools variable entered as a block variable.

Document-based question. Two variables had significant beta values in the total sample when the document-based question (DBQ) was entered into the multiple-regression model as the dependent variable. Sex had a significant beta value as did cognitive orientation (Para. Cog.). For females, PSAT/M and Truth/R showed significant beta values. There were no significant beta values for predictor variables in the male sample (see Table 7).

Hypothesis 14, that truth as relative would be a significant predictor of DBQ scores for females, was supported. Hypothesis 15, that a narrative cognitive orientation would be a significant predictor of DBQ scores for females, was not supported.

R^2 values for two of the three models for predicting DBQ scores were statistically significant. Only the model which used the male subgroup for subjects was not significant. Effect sizes were large for all three DBQ models.

Standard essays. One predictor variable, Bem M, had a significant beta value in the total sample with essays as the dependent variable. For female students, the beta value for PSAT/V was a significant predictor. Bem M showed a significant beta value for males (see Table 7).

Hypothesis 16(a) -- that PSAT/V would be a significant predictor of essay scores for males -- was not supported. Hypothesis 16(b) -- that PSAT/V would be a significant predictor of essay scores for females -- was supported.

R^2 values for the three essays score models using the total sample, and female and male subgroups showed high effect sizes and were statistically significant (see Table 7).

R²-Change Models

The 12 R²-change models, which were created using the same sets of predictor and outcome variable and same groups and subgroups of subjects as the multiple regression models (above), are illustrated in Figures 1-4. Each figure has a different outcome variable (Total Score, MC Score, DBQ Score, and Essay Score) and includes data generated from using three sets of subjects (total sample, female students, and male students). Values for the total sample are found above the line between variables; female and male values, respectively, are immediately below. The R²-change value between Sex and Bem Category was entered for the total sample only.

Exogenous factors and intermediate variables. The amount of variance accounted for between and among the exogenous factors (Sex and School) and intermediate variables (Bem Category, Cognitive Orientation, PSAT/V and PSAT/M, and Historical Truth) was the same across all outcome variables and subject groups (the total sample and the female and male subgroups). Accordingly, with the exception of the four outcome variables (Total Score, M-C Score, DBQ Score, and Essay Score), comments regarding R²-change values and the magnitude of effect sizes are applicable across all 12 models. R²-change values represent squared correlation coefficients where only one variable is used as a predictor (e.g., Bem Category and Cognitive Orientation), additional variance accounted for as the last variable entered into a multiple regression model when there is more than one predictor variable (e.g., School and Bem Category predicting PSAT/V values), and as squared canonical correlation when the outcome variable is treated as a block variable (e.g., when Historical Truth is predicted by Bem Category, Cognitive

Orientation, and School). See Figures 1-4 for the R^2 -change value for all exogenous and intermediate variables (with the exception of relationships with outcome variables).

The effect of Sex on Bem Category was calculated only for the total sample. For this sample (all students in the study) though the effect size was only moderate, Sex was statistically significant in accounting for variance of Bem-category values.

The exogenous variable School was significant in accounting for variance in PSAT/V values across all three groups used in the analysis. School also showed significant R^2 -change values in relation to PSAT/M scores for the total study sample and the female subgroup (School was not a significant predictor for male PSAT/M scores). Effect sizes were large for all subject groups for both PSAT values. Though the effect size was only moderate for total subjects, it was high for both female and male subgroups.

Bem values (considered as a block variable) had a significant R^2 -change value in accounting for variance in PSAT/M scores for the total sample and for males. No such significant relationship was found between Bem values and PSAT/V scores. Bem values were not significant predictors of cognitive orientation for either the total sample, females, or males. Effect sizes were low for all subgroups of subjects. As one of three predictors (along with Cognitive Orientation and School), Bem did not show a significant R^2 -change value in accounting for variance of responses to questions regarding the nature of historical truth. Effect sizes were minimal for all groups of subjects.

Cognitive Orientation was used along with Bem and School to predict responses to questions regarding the nature of historical truth. No significant R^2 -change values were found for any of the groups used for analysis. Effect sizes associated with Cognitive

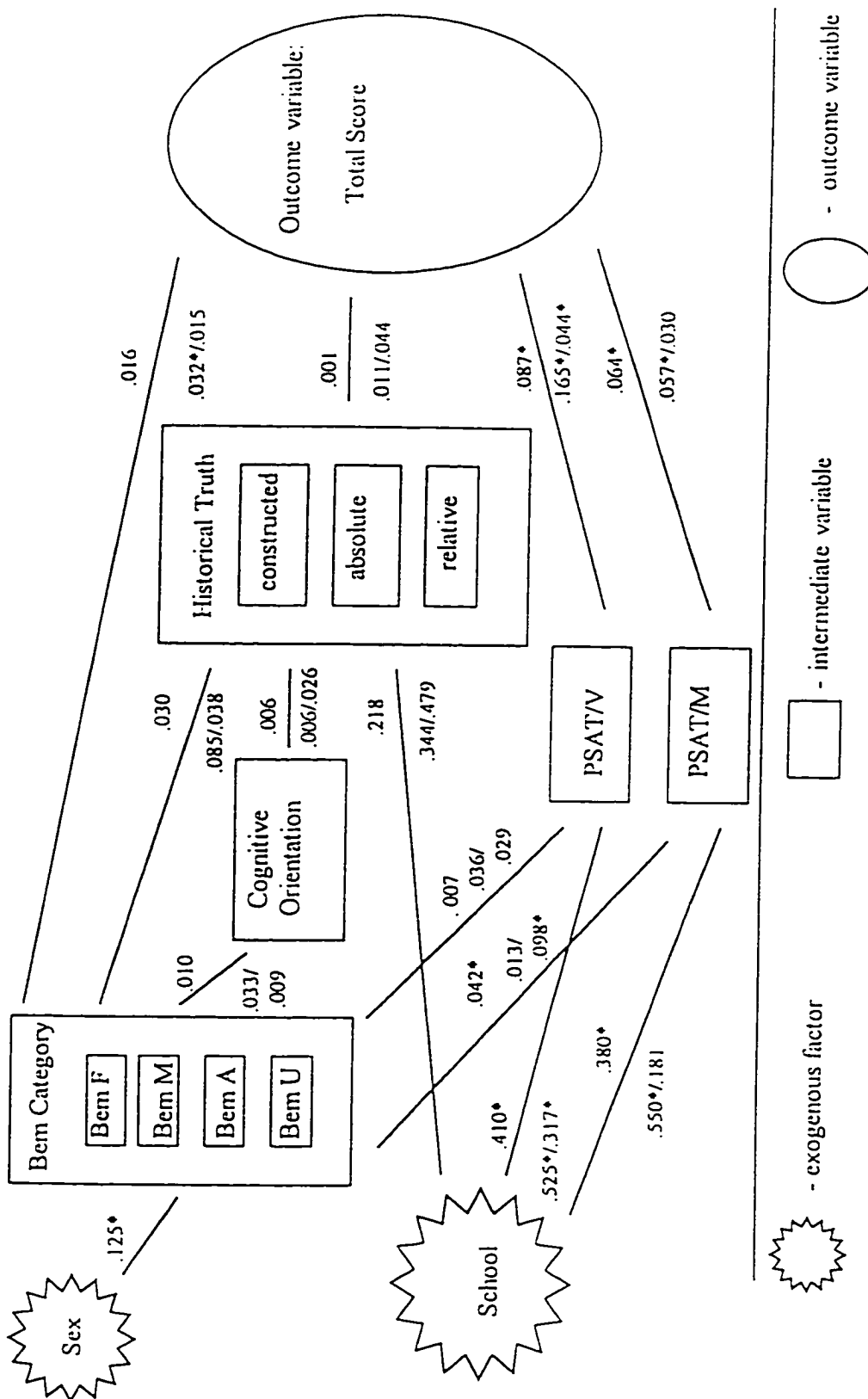


Figure 1. Model representing influence of exogenous and intermediate variables on AP US History achievement. Outcome variable: Total Score

Note. *p<.05. All values are R²-change values. Total-sample value listed first, followed by female/male values.

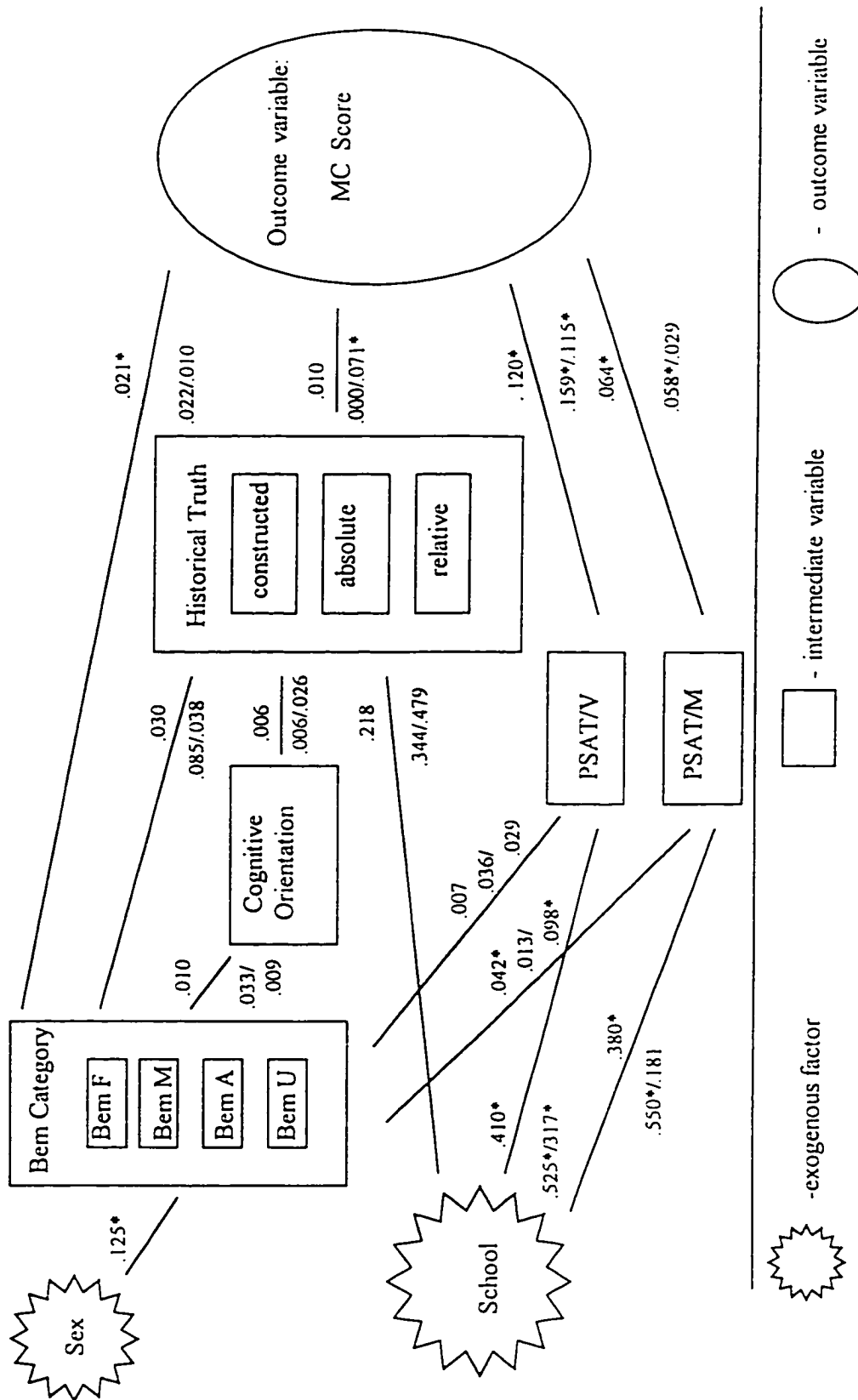


Figure 2. Model representing influence of exogenous and intermediate variables on AP US History achievement. Outcome variable: M-C Score
 Note. * $p < .0$. All values are R²-change values. Total-sample value listed first, followed by female/male values.

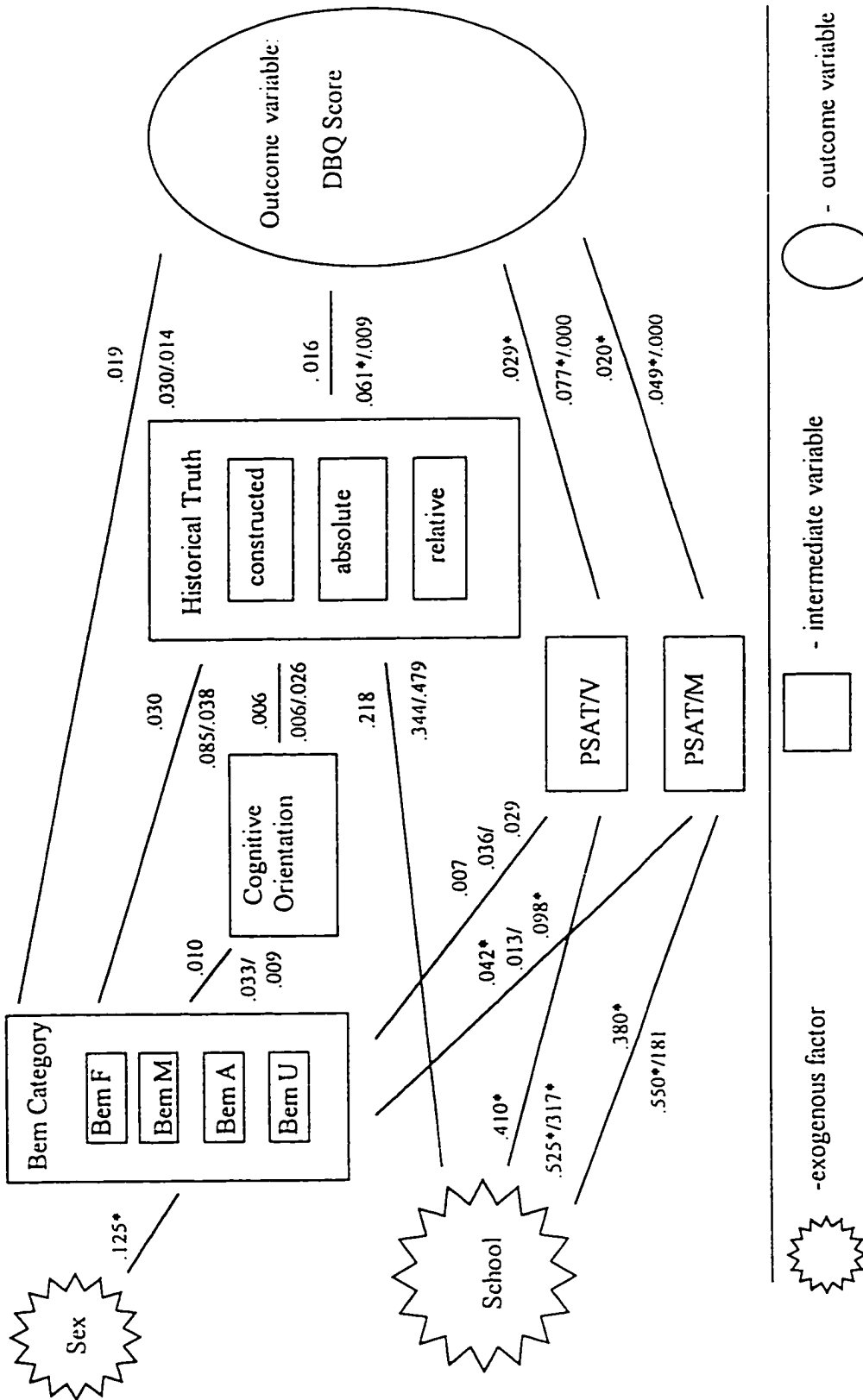


Figure 3. Model representing influence of exogenous and intermediate variables on AP US History achievement.

Outcome variable: DBQ Score

Note. *p<.05. All values are R²-change values. Total-sample value listed first, followed by female/male values.

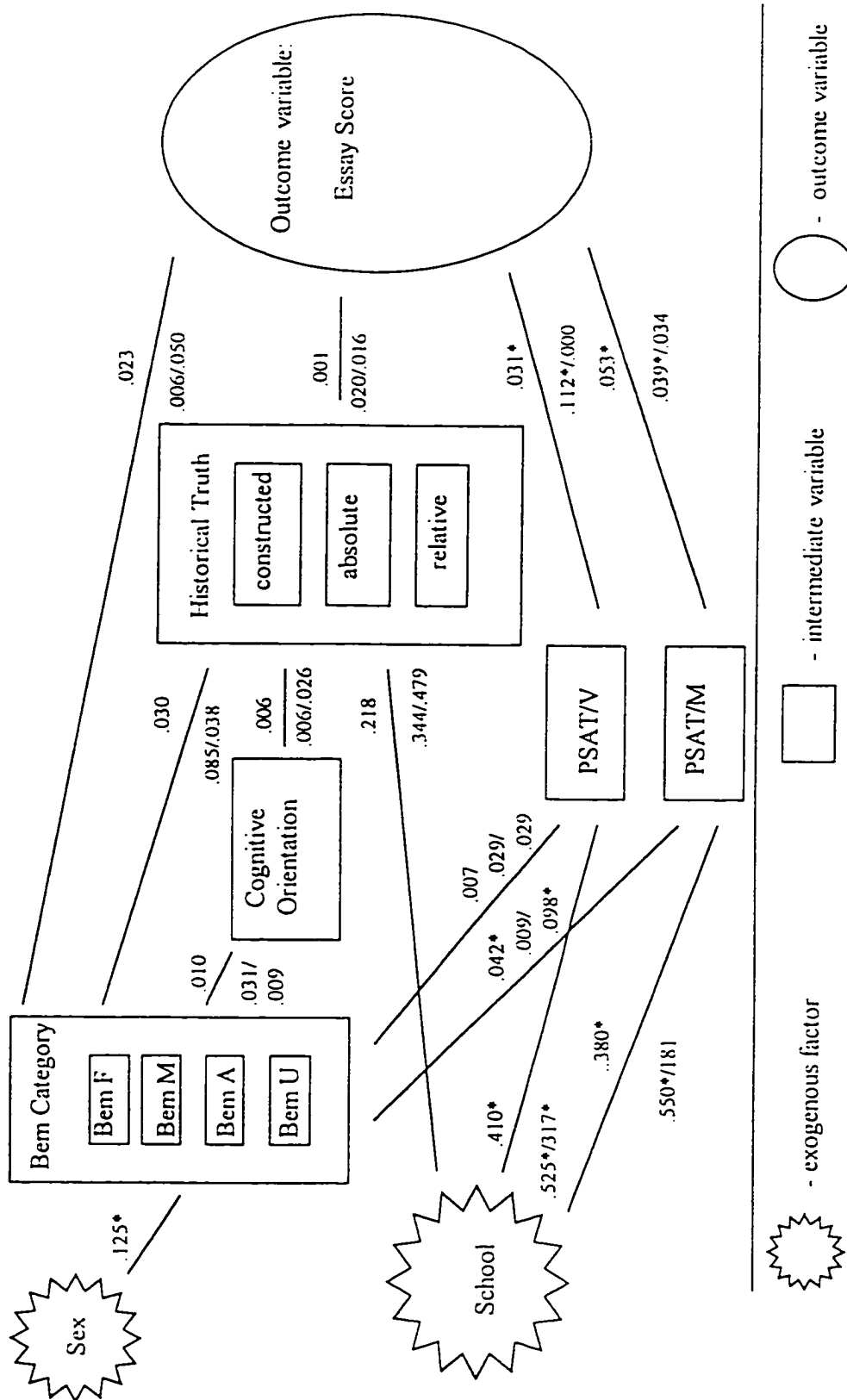


Figure 4. Model representing influence of exogenous and intermediate variables on AP US History achievement. Outcome variable: Essay Score

Note. *p<.05. All values are R²-change values. Total-sample value listed first, followed by female/male values.

Orientation and Historical Truth were minimal.

Bem was not significantly related to Cognitive Orientation for either females or males. Hypotheses 17(a) and 17(b) were therefore not supported. Cognitive Orientation was not significantly related to nature of historical truth for either males or females. Hypotheses 18(a) and 18(b) were therefore not supported.

Total Score. Four variables were included in the models as direct predictors of each of the four outcome variables. For Total Score, Bem resulted in a statistically significant R^2 -change value for the female group of subjects. Effect sizes were small for all subject groups. Historical Truth was not significant for any group of subjects. PSAT/V, on the other hand, was significant for all groups, and showed a moderate effect size for females. PSAT/M had significant R^2 -change values for the total sample and female students. Effect sizes associated with both these groups was small (see Figure 1).

Nature of historical truth did not significantly account for the variance within Total Score for females or males. Hypotheses 19(a) and 19(b) were therefore not supported.

Multiple-choice. For the total sample, Bem showed a significant R^2 -change value. Effect size for the total sample (as well as for the non-significant female and male subgroups) was small (see Figure 2). Historical truth was significant in accounting for variance for the male subgroup, though the effect size for the R^2 -change value was small. PSAT/V was significant for the total sample and both subgroups. Effect sizes for all three groups were in the moderate range. PSAT/M was significant only for the female subgroup. The R^2 -change value associated with PSAT/M and MC Score for females was small.

Document-based question. Bem designations did not show significant R^2 -change values for either the total group or either the female or male subgroups. Effect sizes were small as well for these three groups. Historical Truth was significant for females, though the effect size for the R^2 -change value for females was small. PSAT/V and PSAT/M were both significant in accounting for variance in DBQ score in the total sample and for females. Effect sizes for PSAT scores (V and M) were small.

Standard essays. R^2 -change values associated with Bem outcomes were non-significant for all groups. Effect sizes were small as well. Historical Truth showed a similar pattern -- there was no significant R^2 -change values, and effect sizes were small. PSAT/V and PSAT/M were both significant in accounting for variance in essay scores for the total sample. Effect sizes were in the small to moderate range. For females, PSAT/V showed a significant R^2 value. The effect size for females was moderate. PSAT/M was also significant for females, with the R^2 -change value denoting a small effect size.

Chapter 5

Overview of Findings

For students in the study sample, as for the national population of Advanced Placement US History test-takers in 1997, males outperformed females on the examination as a whole and on the multiple-choice and essay sections of the examination. Females received higher scores than males on the document-based question. The purpose of the study was to explore the possibility that these differences in performance were due to factors other than those traditional speculations offered as explanations for achievement differences in history between males and females -- conventional beliefs that hold males to be somehow better at historical thinking, or that females dislike history as a subject of study and therefore do poorly on assessments of historical knowledge.

When presenting an overview of the research findings, it is reasonable to first consider the most obvious possible explanations for differences in performance between male and female students. These explanations are based on descriptive differences. Following the treatment of descriptive differences, the overview will turn to consideration of findings related to cognitive orientation and the nature of historical truth (truth-in-knowledge), the two major conceptual variables hypothesized to be related in a causal way to performance on the AP US History examination. The overview also includes a consideration of the possible roles played by the school attended and PSAT scores, and concludes with a brief summary of results in the multiple-regression and R^2 -change models.

Descriptive characteristics of students in the sample. To begin the consideration of descriptive characteristics, there is the need to address the question of access to higher-level courses such as AP US History. If academically able girls are somehow selected out (or select themselves out) of courses which require students to perform advanced work, then it would not be surprising if boys outperformed girls on AP tests. The numerical distribution of female and male test-takers nationally and in the study sample speaks to this question of access and selection. Nationally, more females than males take the AP US History examination (53.6% versus 46.4% in 1997). The percentage of females in the study sample is slightly greater -- 55.8% are females, 44.2% males. Though the point could still be made that larger numbers of female test-takers does not preclude the possibility that academically able girls are nevertheless underrepresented in AP US History classes, conversations with some AP teachers in the sample indicated that high achieving girls may actually have been over-represented in their classes, with teachers reporting difficulty convincing boys to take the AP US History course.

Data collected for the study supports the contention that girls in AP US History classes are as able as male students. In the sample, females and males were similar in their past academic achievement. Students of both sexes reported high overall GPAs and high anticipated grades in their AP US History course. PSAT scores were high for both males and females (verbal scores were higher for females and mathematics scores better for males), with scores for both sexes on the verbal and mathematics tests exceeding national means by more than one standard deviation.

Moving from ability and achievement to disposition and motivation, responses to questions that asked about interest in history as a discipline of study also showed little difference by sex. Females chose the least favorable rating of history in comparison to other high school courses more often than males (though they were slightly more likely than males to chose the most favorable rating as well). There was no difference between females and males in the perceived importance of history in forming a world-view and in the likelihood of taking additional courses in history. In addition, though not strictly related to either dispositional or achievement factors, there was a difference in the reported quality of preparation for the AP examination, with males indicating they received better preparation than females.

Taken in combination, these descriptive indicators suggest that both male and female students in the sample were academically able, similar in their past academic achievement, and positively disposed toward the study of history. It is not likely, therefore, that differences in achievement on the AP US History examination between females and males can be ascribed to differences in past achievement or ability, perceived importance of the subject matter, or interest in history as indicated by willingness to continue study in the discipline.

Cognitive orientation. It was hypothesized that students would differ on cognitive orientation according to sex. Specifically, it was hypothesized that there would be a correlation between cognitive orientation and scores on the multiple-choice (MC) part of the AP US History examination and that cognitive orientation would be a predictor of MC scores for males and document-based scores (DBQ) for females. Finally, in the R^2 -

change models, it was predicted that cognitive orientation would account for significant portions of variance in the nature of historical truth (truth-in-knowledge) for both males and females.

Analysis of the data showed virtually no difference between males and females in their cognitive orientation. Both were coded as giving paradigmatic responses approximately 60% of the time. In addition to calculating and comparing the overall percentage of narrative and paradigmatic responses for each sex, an attempt to assess the consistency of responses was carried out by determining the percentage of students who were coded as providing either narrative or paradigmatic responses 60% or more of the time. Males and females did not differ on this measure of consistency, with 3.6% of the males and 1.9% of the females showing a consistent narrative pattern, and 50.0% of the males and 54.7% females coded as consistently paradigmatic in their responses. No correlation was found to exist between cognitive orientation and multiple-choice scores.

In the multiple-regression models, cognitive orientation was not a significant predictor for males on the multiple-choice section and was not significant in predicting document-based question scores for females (though it was a significant predictor for DBQ scores for the all-subjects group). Interestingly, though (paradigmatic) cognitive orientation was positively related to multiple-choice scores for all three groups (i.e., males, females, and all-subjects) and negatively related to both DBQ and Essay Scores for all groups, with the exception of DBQ/all-subjects, cognitive orientation appears to have little statistical predictive value.

In the R^2 -change models, which examined the extent to which individual variables accounted for variance in other dependent and independent (outcome) variables, cognitive orientation did not account for significant variance in the set of three historical truth variables, the one set of variables that was hypothesized to be directly influenced by cognitive orientation.

Contrary to the hypothesis predicting sex-related differences in cognitive orientation, males and females gave similar numbers of narrative and paradigmatic responses on the Narrative/Paradigmatic Organization of Knowledge questionnaire. Furthermore, students in the study appeared to use both narrative and paradigmatic ways of organizing information. As noted above, these “mixed-orientation” responses were the norm across the students in the sample, with very few students being consistently narrative in their responses and only slightly more than half being consistently paradigmatic.

The similarity in responses for males and females and the pattern of responses observed for all students may be due to a number of factors. One possibility is that the male and female responses simply approximate the way things are; that is, the data could be taken to support the position (unstated in Bruner’s writings but clearly stated in personal conversation) that cognitive orientation does not differ by sex. However, the lack of consistency in responses by both sexes is a problem for Bruner’s idea that individuals have a preferred mode of cognitive orientation. The data clearly did not support this presumption.

Both of these results -- that cognitive orientation does not differ by sex and that individuals do not exhibit a consistent orientation -- may reflect the reality of cognitive

orientation for the population at large. Yet it is also possible that the instrumentation of Bruner's notion of narrative and paradigmatic orientation was itself less than perfect.

More problematic than questions of instrumentation, the lack of correlational and predictive significance regarding cognitive orientation is more troublesome conceptually. Though not necessarily an undesirable characteristic in itself, cognitive orientation appears to be unrelated correlationally to other predictor variables (e.g., Bem, nature of historical truth, PSAT) and only weakly related to the outcome variables of student performance on the AP US History examination. As a predictor, it is again weak, showing significance only in one situation (DBQ scores for the all-student group). It did not account for variance in the assumed relationship with conceptions of historical truth.

It would appear that either the concept has not been properly operationalized (see Limitations of the Study and Recommendations for Further Research) or that cognitive orientation is truly (as I suspect Bruner might claim) an underlying, pre-epistemic factor more on the lines of a learning style which, when taken in combination with subsequent psychological and environmental factors, contributes only modestly to learning and achievement. Finally, in light of the findings from this study, it may also be the case that Bruner's concept of cognitive orientation was tested, with the results suggesting either that such orientation does not exist or that it exists but its expression does not differ by sex or gender.

Nature of historical truth. In addition to the prediction of difference in cognitive orientation for males and females, it was hypothesized that males and females would also differ in their responses to questions concerning the nature of historical truth (truth-in-

knowledge). That is, it was expected that females would show more truth-as-relative responses while males would favor more truth-as-absolute responses. In addition, it was predicted that for both sexes, (a) there would be a significant correlation between historical truth-as-relative scores and performance on the document-based question (DBQ) section of the examination, and (b) there would be a significant correlation between a historical truth-as-absolute orientation and performance of the multiple-choice section of the examination. More specifically, and sharpening the predictions by sex, it was also hypothesized that truth-as-absolute would be a significant predictor of multiple-choice scores for males and truth-as-relative would be a significant predictor of DBQ scores for females in the multiple-regression models. Finally, in the R^2 -change models, it was hypothesized that conceptions about the nature of historical truth would account for a significant portion of the variability in all four of the outcome variables (total examination score, MC score, DBQ score, and Essay score) for both males and females.

The predicted difference in responses between males and females to questions about the nature of historical truth was not found to exist. Students of both sexes showed a tendency to think of historical truth as constructed (students chose this definition of historical truth nearly 45% of the time), and there was virtually no sex difference in the distribution of responses across the three variables of epistemic orientation. A measure of consistency in response was calculated for the nature of historical truth set of variables. A value of 60% or greater was again used as the cutoff to distinguish subjects who were consistent in their epistemic orientation from those who were not. For both sexes and cross the three orientations, 19.4% of the students exhibited consistency in their

responses. Very few students were categorized as consistently absolute in their conception of historical truth (2.4% males and 0.9% females). Slightly more were consistently relative (3.6% males and 5.7% females). Most students who were consistent in their epistemic viewpoint thought of historical truth as constructed, with 9.5% males and 16.0% females being so identified. Because the number of students who were consistent in their epistemic orientation was small, comparisons by sex should be approached carefully -- though it appears that more females may think of historical knowledge as constructed than males. Of more importance is the small number of students -- male or female -- who were classified as consistent in their responses, regardless of the chosen epistemic orientation. Fewer than one in five students indicated a preference for one definition of the nature of historical truth for 60% or more of their responses.

There was no correlation between truth-as-relative values and DBQ scores across the total sample, though truth-as-absolute did show a significant correlation with multiple-choice scores. For the multiple-regression models, truth-as-relative predicted DBQ scores significantly for females, and, showing a negative relationship, truth-as-absolute predicted multiple-choice scores for males.

In the R^2 -change models, the three nature of historical truth variables (taken as a group) followed the pattern shown in the multiple-regression models, accounting for a significant amount of variance for males with regard to the multiple-choice outcome variable, and for females with regard to the document-based question variable.

Though hypotheses relating to conceptions of historical truth were found significant more often than those proposed for cognitive orientation, they did not offer overwhelming support that a particular view of historical truth provides much in the way of explaining performance of the AP US History examination. As for the concept of cognitive orientation, this may be so for a number of reasons.

There may be a problem with assessing the concept of the nature of historical truth. It is possible that the study failed to develop an effective way to measure epistemic orientation. Furthermore, few students were categorized as being consistent in their responses to the Nature of Historical Truth questionnaire. If students hold a singular view of historical truth, it is clear that the instrument used in this study was unable to measure it. On the other hand, it may be that students do not think of historical truth in a consistent manner and that the questionnaire was well constructed and did an acceptable job of identifying which students associated different historical scenarios and questions with particular beliefs about historical truth. See Limitations of the Study and Recommendations for Further Research for further discussion of this question of instrumentation.

The second possible explanation for the findings involving historical truth moves to the realm of theory undergirding the concept of what I am calling epistemic orientation. Such an explanation would call for a consideration of theoretical notions involving gendered socialization and subsequent behavior, specifically the extension of findings in sex-related decision-making behavior to academic settings.

Beginning with the work of Chodorow (and her claims that female children are more likely to be socialized into behavior marked by extensive interpersonal relationships than boys) and running through the work of Gilligan and the women's way of knowing approach (which can be taken broadly to claim an interpersonal orientation that is subjective and situation-based), a case can be made that women operate from a position of humanistic interconnectedness while men are more likely to apply rules of impersonal and general logic.

Yet the results of the study do not indicate a level of robustness that would permit these claims to be extended into the everyday world of learning in high school classrooms. Males and females were virtually identical in their responses to questions about the nature of historical truth. When the nature of historical truth was used to predict achievement on the AP US History examination, there were only two instances where females differed from males -- truth-as-relative responses showed some predictability on document-based question scores for females, and truth-as-absolute responses showed some predictability on multiple-choice scores for males.

Females may construct and navigate through their worlds using guideposts of social interconnectedness and males may be more inclined to use abstract principles of formal logic. These ways of organizing lives and responding to everyday problems may be highly regular and predictable for many types of conventional and novel tasks. Yet it may also be the case that such orientations are only tangentially related to epistemic orientation, at least as epistemic orientation shows itself in high school classrooms. That

is, a person may use one orientation when engaged in cognitive activity such as moral decision-making and another when thinking about the existence of historical truth.

It may also be that the ways of learning and thinking about knowledge differ across academic disciplines regardless of sex. What counts as good thinking in History may differ from thoughtful discourse in English or Biology. Also, ways of thinking about knowledge may differ within a discipline depending on the level of understanding within the discipline. A course taught at an AP (or other advanced) level may contribute to the development of epistemic beliefs differently than non-AP courses.

School and PSAT scores. The study did not include any hypotheses about the relationship between the school attended and performance on the AP US History examination. Similarly, there were no predictions made about the relationship between PSAT scores and examination performance (though PSAT scores were included in the analysis of sex differences in descriptive data). It is useful, however, to look briefly at these variables in the multiple-regression and R^2 -change models and to speculate about the reasons for the significant predictive and explanatory values that are associated with PSAT and School variables.

The variable School was included in both the multiple-regression and R^2 -change models. Though the multiple-regression analysis did not determine the predictive value of the School variable in isolation (School was entered in each multiple-regression model as a block variable and partial correlation coefficients for each school were therefore not computed), since there were so few significant values associated with other variables (with the notable exception of PSAT scores), it is likely that School was an important

contributor to the 11 of 12 models that showed significant overall R^2 values. In the R^2 -change models, the School variable clearly accounted for most of the variance in PSAT scores. In a separate analysis, when School was entered as the first of four blocked variables in multiple-regression models that used total score, multiple-choice score, document-based question score, and essay score as independent variables, it showed R^2 (change) values of .311 to .497, accounting for 80% of the explained variance in three of the models and 70% in the fourth (see Table 8).

PSAT scores were also good predictors of partial and total examination performance, though this was again more the case for females than for males. PSAT/V was a significant predictor for females for the test as a whole, and for both males and females for items presented in a multiple-choice format. It was also a significant predictor for females on the essay questions. PSAT/M was significant for females for the test as a whole, as well as being a significant predictor for questions presented in multiple-choice and document-based formats. In addition, the PSAT variable, though producing relatively small effect-sizes, showed the highest R^2 -change values in the set of four variables used to predict overall and partial scores on the AP examination in the R^2 -change models and accounted for a significant amount of variance for each of the outcome variables, particularly for females (the percent of variance accounted for by PSAT/V on Total Score was four times greater for females than males).

Though other socioeconomic and cultural influences were not considered in the study, the importance of factors such as family income and educational level of parents in relation to student achievement on standardized tests such as the PSAT has long been

Table 8

R² and R²-Change Values for Four Multiple-Regression Models
Across Four Outcome Variables

Model	Total Score		Multiple-Choice		Document-Based		Essays	
	R ²	R ² -change	R ²	R ² -change	R ²	R ² -change	R ²	R ² -change
1	.497*	-	.475*	-	.311*	-	.328*	-
2	.575*	.077	.632*	.157	.312*	.001	.371*	.043
3	.606*	.031	.667*	.035	.348*	.036	.409*	.038
4	.609*	.003	.679*	.012	.384*	.036	.411*	.002

Note. *p<.05. R² values are for overall model. Model 1 has Schools as a block independent variable. Model 2 adds PSAT/V and PSAT/M as second block. Model three adds Bem, Nature of Truth, and Cognitive Orientation as third block. Model 4 adds Sex as fourth blocked independent variable. Data for all students used for each model.

documented. For students who attend public schools, there is often a stratification of families and students along income, educational, and class lines. Although these socioeconomic variables are not likely to map exactly onto the factors that ultimately account for achievement (the relationship between and among socioeconomic factors and ability is, of course, still an open question), it should not be surprising that students from communities where family incomes and levels of educational attainment are high do well on PSAT tests, while students in schools with larger numbers of students from lower income families whose parents have less educational attainment receive lower scores. And though scores on AP examinations to a great extent reflect the quality of instruction in a single course (as opposed to the abilities assumed to develop over time that are necessary to insure high scores on the PSAT), it is again not unreasonable to think that students coming to the AP experience with the support and expectations of family and teachers, along with high quality instruction in courses leading to the AP level, will do well in an AP course and receive a high score on an AP examination.

There would seem to be at least two reasons why PSAT performance is so closely related to achievement on the AP US History examination. The first reason is correlational. The second is more likely to be causal in nature. The first reason involves the similar item types in both the PSAT and AP US History examination. PSAT tests are made up of questions administered in the same (multiple-choice) format as the majority of questions that comprise the AP US History examination. If there were factors contributing to high performance on multiple-choice questions, particularly questions with a broad verbal domain as content, it would be reasonable to find that students perform

similarly on the two different tests. In addition to similarity in item-type, it may be that the PSAT tests assess, as the College Board and ETS claim they do, academic skills that enable students to master academically demanding coursework. If this were the case, it would again be reasonable to suppose that students with such strong general academic readiness would be better able to learn the content in a college-level course such as AP US History and perform well on the AP examination.

R² in multiple-regression and R²-change models. Because much of the information contained in the multiple-regression and R²-change models has been noted above (see particularly the important effects of School and PSAT scores immediately above) , only a brief summary of the findings which emerged from analysis of the models will be presented in this part of the summary.

With one exception, the 12 multiple-regression models produced high R² values; that is, when Sex (used for the total sample only), Bem Category, PSAT scores, nature of historical truth values, and cognitive orientation were use to predict total and part scores on the AP examination in US History, the models accounted for substantial variance in the outcome variable in 11 of 12 models. The exception occurred when the document-based question (DBQ) was the outcome variable and males were the subject group. It should also be noted that in each of the 12 models, the School variable was included in the analysis and may well have accounted for a large portion of the explained variance in the multiple-regression models.

The R²-change models were created to examine the relative value of each of the variables considered to be either totally independent (exogenous) or intermediate-level on

other specific intermediate variables and four outcome variables (Total Score, Multiple-Choice Score, DBQ Score, and Essay Score). In these models, only the School variable and PSAT scores were consistently high in accounting for variance in other variables, though Bem Category contributed to explaining variance in PSAT/M scores for males and School accounted for much of the variance in overall PSAT scores. PSAT scores, in turn, accounted for much of the variance in the four outcome variables.

Moving temporally within the models (going from left to right reflects earlier to later periods in time), the overall R^2 -change model design did not show Bem Category to account for significant amounts of variance in Cognitive Orientation, nor were variances in Historical Truth accounted for by either Bem Category, Cognitive Orientation, or School.

In the R^2 -change models, four intermediate variables (Bem Category, Historical Truth, PSAT/V, and PSAT/M) were hypothesized to directly account for the variance in the four outcome variables. The effect of the first two variables was limited. Bem Category accounted for significant variation in Total Score for female students, and for all students when Multiple-Choice Score was the outcome variable. Historical Truth was significant in accounting for variance for multiple-choice scores for males and for DBQ scores for females.

Overview summary. What emerges from the data is a pattern that shows a group of academically able students who have a history of past academic achievement, and who are positively disposed to learn history. This characterization holds for both male and female students. And regardless of the school they attend (a factor admittedly of great

importance in the search for determining examination performance) males and females are, nevertheless, much more alike than different in their academic and attitudinal profile.

Multiple-regression and R^2 -change models show that PSAT scores and the school attended are good predictors of scores on all parts of the AP US History examination. PSAT scores predict particularly well for females. The developed academic competence which is claimed to be measured by the PSAT (particularly the verbal part of the PSAT) is much more important in accounting for total and part-score performance on the AP US History examination than measures of gender, cognitive orientation, or beliefs about the nature of historical truth.

Limitations of the Study

The process of constructing the sample was carried out without difficulty and the sample was representative of public high schools in New England which offer AP courses in US History. Very few students chose not to participate in the study. Forty students were excluded because they were enrolled in grade 12. The decision was made to include only students who responded to a majority of items on each of the four research instruments (with PSAT scores a required element on the Student Biographical Questionnaire). This resulted in 53 students being excluded from the dataset (mostly due to missing self-reported PSAT scores). The final sample size was 190. It is possible that the excluded students were on the whole less able than those included in the study sample, though a comparison of ability and achievement measures for the two groups yields mixed results. Excluded students performed less well on the examination than students in the

sample ($t=2.138$, $df\ 241$, $p=.034$), yet there was no difference between the two groups on PSAT scores and estimated course grade.

Inspection of responses to questions on the Bem Sex-Role Inventory, the Nature of Historical Explanation and Truth, and Narrative/Paradigmatic Organization of Knowledge questionnaires showed a lack of straight-line responding (i.e., choosing the same answer to all questions), indicating active engagement of students in completing the three questionnaires used in the study. As has been noted, however, such variability in student responses has important theoretical implications. Because students did not produce consistent responses to questions about their conceptions of the nature of historical truth or cognitive orientation, additional work may need to be carried out to confirm the existence of these two constructs.

Because both of the major intermediate variables were constructs for which no measurement instrument existed prior to the study (the Nature of Historical Explanation and Truth and the Narrative/Paradigmatic Organization of Knowledge questionnaires were developed for the study), it is possible that measurement error was present. Specifically, while inter-rater reliability for the assessment of cognitive orientation (which was assessed by using nine open-ended questions) was satisfactory, it may be that the use of an open-ended instrument allowed students to respond more with socially-learned answers rather than actually assessing their cognitive preference in organizing new information. To determine the effect of the manner of assessment, a version of the survey with different item types could be developed and administered to a comparable sample of students.

Recommendations for Further Research

After considering the major findings of the study, three areas emerge as candidates for further research. The first involves further attempts to operationalize the concept of cognitive orientation. The second would extend the suggestive findings of advanced epistemological development of students enrolled in AP US History courses. The third concerns the need to continue research into the influence of the school and school climate on achievement.

Cognitive orientation. Teachers as well as psychometricians know that there is a difference between assessments that ask students to create a response and those which require choosing among responses generated by others. In assessment, creating an answer is often seen as better -- as requiring more complex mental skills -- than choosing between existing answers. It was in this sense of asking subjects to generate their responses that the instrument to assess cognitive orientation for this study was developed.

Yet it may be that cognitive orientation is not best assessed by means of a creative, open-ended activity, particularly if the task allows for the inclusion of learned knowledge that itself may be considered narrative or paradigmatic. As an example, consider the first scenario in the Narrative/Paradigmatic Orientation of Knowledge questionnaire. The task required students to think of questions to ask when inquiring about possibly buying a car that was advertised in a newspaper. Because of a limited range of learned behavior (e.g., a societally-reinforced script that includes an overriding concern with price) a student's dominant cognitive orientation might be masked by previously learned socially-dominant responses. Before accepting the position that there is no difference in cognitive

orientation between females and males, additional research could be carried out to identify novel scenarios where prior experience is unlikely to influence the choice between equally desirable narrative and paradigmatic responses to organizing information. From this, an instrument consisting of these forced-choice questions could be constructed to use in an assessment of cognitive preference by sex.

In addition, the questionnaire used in the study was designed to elicit responses that reflected cognitive preference in general lifestyle situations for high school students. A pattern of the sort shown by students in the study might differ if the content of the questions or the age or other characteristics of the subjects is different. Additional work should be done to examine the relationship of cognitive preference employed in conjunction with a variety of informational content for individuals who differ by age and other demographic characteristics.

Epistemological development. Students in the study showed a level of epistemic sophistication far beyond what current theory on epistemological development would predict. It is difficult, of course, to know the exact causes of such development -- after all, AP students are likely to have a history of high academic achievement long before they take an AP course. Regardless, students in the study were classified as consistently constructive in their thinking about the nature of historical truth almost three times as often as they were classified as consistently relative (less than 2% of the students were considered consistently absolutist in their orientation toward historical truth). With respect to the total set of approximately 1,700 responses given to the nine historical

scenarios and accompanying questions, the constructed sense of historical truth was chosen nearly 45% of the time.

From Perry through Baxter-Magolda and King and Kitchener, researchers have found that first-year college students almost always thought of knowledge as definite and knowable and unchanging. This certainty of knowledge belief was followed by a period of relativism developed during the second and third years of college, which was then eclipsed toward the conclusion of the undergraduate experience by a realization of the socially-constructed nature of much knowledge and truth. When considered this way, the students in the study showed a level of epistemological sophistication not expected to emerge until nearly the end of a traditional collegiate education.

The AP curriculum in US History and the pedagogical activity associated with it may be contributing to this advanced epistemic status in students. It is not simply that students who take AP courses are cognitively superior (after all, Perry was using Harvard undergraduates in his work). Rather, it may be that it is possible to model more sophisticated thinking about events and issues in a course like AP US History, teaching the course in ways that focus on the analytic interpretation of information with a skepticism toward definite answers of historical causality. As a result of such a pedagogical approach, it is not unreasonable to think that students would show a shift toward a more generalized constructivist view of historical knowledge and truth. Additional research could be carried out to investigate the extent to which students showed similar epistemic development in non-AP level history courses, what their patterns of epistemic characterizations were in history courses taken at lower grade levels, and,

though it would involve construction of somewhat different conceptual categories (e.g., what counts as a constructed truth in English), to carry out similar assessments in courses from academic disciplines other than history.

In addition, a more complete analysis of the data gathered for the study could provide additional information about the extent and nature of students' ideas about the nature of historical truth. For example, the relationship between prior achievement (using GPA and PSAT scores) and the pattern of students' selection among the three variables of the nature of historical truth could be examined.

School and school climate. The present study was designed to assess the importance of two cognitive variables in determining AP examination performance. The data showed only small effects for these variables. On the other hand, though not a direct research interest in the study, it is clear that academic achievement (at least in the sense that AP scores are considered an indicator of academic achievement) is very much related to the school a student attends. The R^2 -change values show the School variable to be significant in all but one condition in predicting PSAT scores. The School variable accounted for approximately four-fifths of the variance in the total score on the AP US History examination and for the three parts of the examination.

PSAT scores are, in turn, significant predictors of performance in all but seven of the 24 possible relations between PSAT score and examination performance. These high prediction values do not appear for either the cognitive variables (cognitive orientation and nature of historical truth) or sex and gender measures. In addition, the Schools

variable was a strong predictor of achievement (nearly half of the variance in total score performance for all students was explained by Schools).

Building on these suggestive findings, and using data already available from the study, further research could explore the cluster of descriptors associated with high levels of performance in schools at which students did well on the examination. For example, it may be that schools offering large numbers of AP courses and having large numbers of students taking AP examinations have better results than schools with smaller AP populations. Demographic information about teachers who teach AP courses (are AP teachers more experienced than the faculty as a whole?) would be useful to include in models involving factors that influence achievement across schools, as would data related to direct teacher involvement with AP (e.g., attendance at AP workshops and Institutes, serving as readers of AP examinations).

Measures of family income and other socioeconomic indicators (e.g., education level of parents, occupational classification) could be used as predictor variables in models developed to account for AP scores across schools. The effect of broader community involvement with AP is another area of interest. Does the school district require the high school to offer AP courses? Does the school board pay for all or part of AP examinations in the district? Is AP limited to students with extremely high past academic performance, or are other criteria used to recruit students into AP courses? Other demographic and achievement data from state-level sources (e.g., per pupil expenditures, direct financial support for advanced classes such as AP) could be gathered to create an even richer descriptive context.

A related line of research that incorporated qualitative techniques could investigate the direct influence of the teacher on achievement. Interviews with students and teachers could be done with the goal of developing a profile of successful AP US History teachers and schools, the profiles then being matched with the cognitive characteristics of successful (and less successful) students.

Conclusion

The question of why males receive better scores on the AP US History examination than females served as the starting point for this study. In the course of trying to answer this question, the research focus has moved from descriptive (males do better primarily because they receive higher scores on a heavily-weighted section of the examination that is administered in multiple-choice format) to conceptual-level explanations. It was hypothesized that differences between males and females in characteristic ways of processing information and in thinking about the nature of historical truth would account for differences in performance on questions administered in different item-formats.

For the most part, the hypothesized differences by sex in cognitive and epistemic orientation were not supported by the research data. Though many students of both sexes appeared to think about historical truth in a sophisticated manner, in a predictive sense it was more important to know a student's PSAT scores and the school s/he attended than it was to know about how she processed new information or whether he considered historical truth to be known with certainty, to be completely relative, or to be socially constructed.

In a way, the data have led full circle. The R^2 values for the multiple-regression models and the R^2 -change values for individual variables in the path models showed that a relatively small number of variables taken together could account for much of the differences in performance on the AP US History examination. But the outcome was once again more descriptive than explanatory. The data suggest the importance of analyzing environmental factors such as the school attended when trying to understand differences in performance. Yet even if school-related factors were found to play a significant role in determining performance, it would still be necessary to explicate how such factors were translated from outside influences to attitudes and knowledge within an individual student and how the resulting dispositions and subsequent learning were demonstrated on instruments of assessment.

It would appear that the study's original question regarding the reasons for sex-related performance differences on the AP US History test continues to exist. To insure fairness in opportunity for all students, regardless of sex or gender, the search for the reasons needs to continue as well.

References

- Agruso, S. A. (1990, April). Performance of Ninth Grade Males and Females on a Minimum Competency Test in Science. Paper presented at the annual meeting of the American Educational Research Association, Boston, MA.
- Atkins, W. J., Leder, G. C., O'Halloran, P. J., Pollard, G. H., & Taylor, P. (1991). Measuring risk taking. Educational Studies in Mathematics, 22, 297-308.
- Baxter-Magolda, M. B. (1992). Knowing and reasoning in college. San Francisco: Jossey-Bass.
- Beal, C. R. (1994). Boys and girls: The development of gender roles. New York: McGraw-Hill.
- Belenky, M. F., Clinchy, B. M., Goldberger, N. R., & Tarule, J. M. (1986). Women's ways of knowing. New York: Basic Books.
- Bell, R. C., & Hay, J. A. (1987). Differences and biases in English language examination formats. British Journal of Educational Psychology, 57 (2), 212-220.
- Ben-Shakhar, G., & Sinai, Y. (1991). Gender differences in multiple-choice tests: The role of differential guessing tendencies. Journal of Educational Measurement, 28 (1), 23-35.
- Bem, S. L. (1981). Bem sex-role inventory. Palo Alto: Mind Garden.
- Bennett, C. K., & Bennett, J. A. (1994, April). Teachers' Attributions and Beliefs in Relation to Gender and Success of Students. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.
- Best, R. (1983). We've all got scars. Bloomington: Indiana University Press.
- Block, J. H. (1983). Differential premises arising from differential socialization of the sexes: Some conjectures. Child Development, 54, 1335-1354.
- Bolger, N., & Kellaghan, T. (1990). Method of measurement and gender differences in scholastic achievement. Journal of Educational Measurement, 27 (2), 165-174.
- Bower, T. G. R. (1982). Development in infancy. (2nd ed.). San Francisco: Freeman.

Breland, H. M., Danos, D. O., Kubota, M. Y., & Sudlow, M. W. (1991). A study of gender and performance on Advanced Placement history examinations (College Board Rep. No. 91-4). Princeton, NJ: Educational Testing Service.

Bridgeman, B., & Morgan, R. (1993). Consequences of discrepancies between performance on multiple-choice and essay questions. Princeton: Educational Testing Service.

Bruner, J. (1986). Actual minds, possible worlds. Cambridge, MA: Harvard University Press.

Bruner, J. (1985). Narrative and paradigmatic modes of thought. In E. Eisner (Ed.), Learning and teaching the ways of knowing. Chicago: National Society for the Study of Education.

Bussey, K., & Perry, D. G. (1982). Same-sex imitation? Sex Roles, 8, 773-784.

Cherry, M. J. (1994). Gilligan's "different voice" theory: A just cause for composition inquiry or just another conceptual bandwagon? Annual Meeting of the Conference on College Composition and Communication, Nashville, TN.

Chodorow, N. (1974). Family structure and feminine personality. In M. Z. Rosaldo & L. Lamphere (Eds.), Woman, culture, and society (pp. 43-66). Palo Alto: Stanford University Press.

Chodorow, N. (1989). Feminism and psychoanalytic theory. New Haven: Yale University Press.

Doolittle, A. E. (1986, April). Gender-Based Differential Item Performance in Mathematics Achievement Items. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Eagly, A. H. (1995). The science and politics of comparing women and men. American Psychologist, 50, 145-158.

Fagot, B. I., & Hagen, R. (1991). Observation of parent reactions to sex-stereotyped behaviors: Age and sex effects. Child Development, 62, 617-628.

Fagot, B. I., Hagan, R., Leinbach, M. D., & Kronsberg, S. (1985). Differential reactions to assertive and communicative acts of toddlers boys and girls. Child Development, 62, 1499-1505.

Fairweather, H. (1976). Sex differences in cognition. Cognition, 4, 231-280.

Feldman, B. I., & Rafferty, E. R. (1993, April). Familial Configuration Influences on the Educational Performance of Urban High School Freshmen. Paper presented at the annual meeting of the American Educational Research Association, Atlanta, GA.

Ferber, M. A., Birnbaum, B. G., & Green, C. A. (1983). Gender differences in economic knowledge: A reevaluation of the evidence. The Journal of Economic Education, (Spring), 24-37.

Flanagan, O. & K. Jackson (1987). Justice, care, and gender: The Kohlberg-Gilligan debate revisited. In M. J. Larrabee (Ed.), An ethic of care: Feminist and interdisciplinary perspectives (pp. 69-84). New York: Routledge.

Frey, K. S., & Ruble, D. N. (1992). Gender constancy and the "cost" of sex-typed behavior: A test of the conflict hypothesis. Developmental Psychology, 28, 714-721.

Friedman, W. J., Robinson, A. B., & Friedman, B. L. (1987). Sex differences in moral judgments? A test of Gilligan's theory. Psychology of Women Quarterly, 11, 37-46.

Garmon, L. C. (1996). Gender differences in stage and expression of moral judgment. Merrill-Palmer Quarterly 42(n3), 418-437.

Gilligan, C. (1982/1993). In a different voice. Cambridge, MA: Harvard University Press.

Gilligan, C., Ward, J. V., & Taylor, J. M. (Eds.). (1988). Mapping the moral domain: A contribution of women's thinking to psychological theory and education. Cambridge, MA: Harvard University Press.

Goldberger, N. (1996). Cultural imperatives and diversity in ways of knowing. In N. L. Goldberger, J. M. Tarule, B. McV. Clinchy, & M. F. Belenky (Eds.), Knowledge, difference and power (pp. 335-372). New York: Basic Books.

Hadfield, O. D., & Maddux, C. D. (1988). Cognitive style and mathematics anxiety among high school students. Psychology in the Schools, 25, 75-83.

Hahn, C. L. (1982). Are there gender differences in high school economics in the eighties? Journal of Economics Education, 13, 57-65.

Halpern, D. F. (1992). Sex differences in cognitive abilities. (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Hassman, P., & Hunt, D. P. (1994). Human self-assessment in multiple-choice testing. Journal of Educational Measurement, 31 (2), 149-160.

Hepburn, E. R. (1994). Women and ethics: A "seeing" justice? Journal of Moral Education, 23, 27-38.

Hudson, L., & Jacot, B. (1991). The way men think: intellect, intimacy, and the erotic imagination. New Haven: Yale University Press.

Huston, A. C. (1983). Sex typing. In E. M. Hetherington (Ed.), Handbook of child psychology: Socialization, personality, and social development (Vol. 4, pp. 387-467). New York: Wiley.

Hyde, J. (1981). How large are cognitive gender differences? American Psychologist, 36 (8), 892-901.

Hyde, J., & Linn, M. (1988). Gender differences in verbal ability: A meta-analysis. Psychological Bulletin, 104 (1), 53-69.

Keefer, M. W. (1993). Gilligan's moral orientation hypothesis: Strategies of justification and practical deliberation. Annual Meeting of the Society for Research in Child Development, New Orleans, LA.

King, P. M. and K. S. Kitchener (1994). Developing reflective judgment. San Francisco: Jossey-Bass Publishers.

Kohlberg, L. (1966). A cognitive-developmental analysis of children's sex role concepts and attitudes. In E. E. Maccoby (Ed.), The development of sex differences (pp. 82-173). Palo Alto: Stanford University Press.

Kornblith, H. (1994). Naturalizing epistemology. Cambridge, MA: MIT Press.

Kuhn, D., Nash, S. C., & Bruken, L. (1978). Sex role concepts of two and three year olds. Child Development, 49, 445-451.

Luria, A. R. (1976). Cognitive development: Its cultural and social foundations (Martin Lopez-Morillas & Lynn Solotaroff, Trans.). Cambridge, MA: Harvard University Press.

Lyons, N. P. (1983). Two Perspectives: On self, relations, and morality. Harvard Educational Review, 2, 125-145.

Maccoby, E. E., & Jacklin, C. N. (1974). The psychology of sex differences. Stanford, CA: Stanford University Press.

Martin, C. L., & Halverson, C. F. (1981). A schematic processing model of sex typing and stereotyping in children. Child Development, 52, 1119-1134.

Martin, C. L., & Wood, C. H. (1987, April). Children's sex-typed interest attributions. Paper presented at the Society for Research in Child Development, Baltimore.

Mazzeo, J., Schmitt, A. P., & Bleistein, C. A. (1993). Sex-Related Performance Differences on Constructed-Response and Multiple-Choice Sections of Advanced Placement Examinations (92-7). New York: College Entrance Examination Board.

Moody, J. D., & Gifford, V. (1987, November). The Effect of Locus of Control on Achievement in Chemistry. Paper presented at the annual meeting of the Mid-South Educational Research Association, Mobile, AL.

Mulkey, L. M., & Morton, P. J. (1991, April). Sex Differences in Verbal and Science Performance for Students from One-Parent Families: Further Tests of Economic and Behavioral Explanations of a Small Effect. Paper presented at the annual meeting of the Eastern Sociological Society, Providence, RI.

Mullen, M. K. (1990). Children's classifications of nature and artifact pictures into female and male categories. Sex Roles, 23, 577-587.

Murphy, R. J. L. (1982). Sex differences in objective test performance. British Journal of Educational Psychology, 52, 213-219.

National Assessment of Educational Progress. (1990). The US history report card (19-H-01). National Center for Education Statistics. Princeton, NJ: Educational Testing Service.

Noddings, N. (1984). Caring. Berkeley: University of California Press.

Paulson, S. E. (1992, March). Adolescents' and Their Parents' Perceptions of Parenting Style: Relations with Achievement. Paper presented at the meeting of the Society for Research on Adolescence, Washington, DC.

Paulson, S. E. (1994). Parenting style and parental involvement: Relations with adolescent achievement. Midwestern Educational Researcher, 7, 6-11.

Payne, O. L. (1992, April). The Effects of Motivation and Classroom Learning Environment on Black Secondary Students' Verbal and Mathematics SAT. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Perry, W. G. (1970). Forms of intellectual and ethical development in the college years. New York: Holt, Rinehart & Winston.

Perry, W. G. (1981). Cognitive and ethical growth: The making of meaning. In A. Chickering (Ed.), The modern american college (pp. 76-116). San Francisco: Jossey-Bass.

Reap, M. A., & Cavallo, A. L. (1992, March). Students' Meaningful Understanding of Science Concepts: Gender Differences. Paper presented at the annual conference of the National Association for Research in Science Teaching, Boston, MA.

Roberts, L. R. (1986, March). Gender Differences in Patterns of Achievement and Adjustment during Early Adolescence. Paper presented at the biennial meeting of the Society for Research on Adolescence, Madison, WI.

Robison-Awana, R., Kehle, T. J., & Jenson, W. R. (1986). But what about smart girls? Adolescent self-esteem and sex role perceptions as a function of academic achievement. Journal of Cognitive Psychology, 78, 179-183.

Robinson, J. P., Shaver, P. R., & Wrightsman, L. S. (Eds.). (1991). Measures of personality and social psychological attitudes. San Diego: Academic Press.

Rubin, D. (1993). Gender influences: Reading student texts. Carbondale, IL: Southern Illinois University Press.

Slaby, R. C., & Frey, K. S. (1975). Development of gender constancy and selective attention to same sex models. Child Development, 46, 849-856.

Smith, T. E. (1992). Gender differences in the scientific achievement of adolescents: Effects of age and parental separation. Social Forces, 71, 469-484.

Smith, J., & Russell, G. (1984). Why do males and females differ? Children's beliefs about sex differences. Sex Roles, 11, 1111-1120.

Spence, J. T., & Helmreich, R. L. (1981). Androgyny versus gender schema: A comment on Bem's gender schema theory. Psychological Review, 88, 365-368.

Stiller, N. J., & Forrest, L. (1990). An extension of Gilligan and Lyon's investigation of morality: Gender differences in college students. Journal of College Student Development, 31, 54-63.

Stipek, D. J., & Gralinski, J. H. (1991). Gender differences in children's achievement-related beliefs and emotional responses to success and failure in mathematics. Journal of Educational Psychology, 83, 361-371.

Vreeke, G. J. (1991). Gilligan on justice and care: Two interpretations. Journal of Moral Education, 20 (n1), 33-46.

Vygotsky, L. S. (1978). Mind in society: The development of higher psychological processes (Michael Cole, Alexander R. Luria, Martin Lopex-Morillas, & James Wertsch, Trans.). Cambridge, MA: Harvard University Press.

Weitzman, N., Birns, B., & Friend, R. (1985). Traditional and nontraditional mothers' communication with their daughters and sons. Child Development, *56*, 894-898.

Wertsch, J. V. (1985). Vygotsky and the social formation of mind. Cambridge, MA: Harvard University Press.

Wetzel, F., & Barton, R. (1995, March). Gender differences in achievement on the Advanced Placement U.S. History Examination: Evidence of a sex-specific approach to learning? Paper presented at the meeting of the Eastern Educational Research Association, Hilton Head, SC.

Williams, J. E. (1993, January). Nonacademic Self-Concept and Gender and Achievement Predictors. Paper presented at the annual meeting of the American Educational Research Association, Atlanta, GA.

Appendix A

Historical Explanation and Historical Truth

Historians do many things: they ask questions about the past and think about what information they need to answer their questions; they develop possible explanations for why something happened and consider the reliability of their primary and secondary sources. Eventually, the work of the practicing historian may result in the development of what can be called *historical explanation*. One way to view historical explanation is to say it is information that does a good job of analyzing the causes and consequences of events that comprise historical situations.

However, even if a historian does a good job of historical explanation, we can still ask if the historian has uncovered the *truth* about that situation. Stated another way, does historical explanation imply that historical truth has also been found? Historians and others interested in history continue to disagree about the answer to this question. Here are three possible ways of answering the question:

A) *Historical explanation does not include the idea of historical truth.*

All historians operate from a subjective or biased perspective (in part because they live in a particular place, time, and culture). Because there is no agreed-upon way to judge the correctness of any historical explanation, competing explanations offered by different historians about the same event or issue are equally valid (or invalid) insofar as historical truth is concerned.

B) *Historical explanation includes the idea of historical truth, but what is considered historically true may change over time.* While the work of historians is subjective (and therefore possibly biased), at any given time one explanation for a historical event can be judged better than another by historical experts. Therefore, historical explanations of an event could be considered true, though at a later date a different explanation could prove to be superior and more truthful.

NOTE TO USERS

Page(s) missing in number only; text follows. Microfilmed as received.

147

UMI

C) *Historical explanation includes the idea of historical truth knowable with certainty.* Historical explanations can be identified as either true or false depending on the quality of the evidence presented. While the true explanation for any historical event may not be known now, it is possible to find it in the future. If two historians disagree about the explanation of an event, one will be proven wrong as evidence is evaluated or additional evidence becomes available.

Now that you have read this introduction to historical explanation and historical truth, please read the nine historical situations and the questions that follow each situation. Assume you had the needed research training and could spend much time and resources researching each question. Given these assumptions, your task is to choose the description of historical explanation and truth (A, B, or C) that would most likely characterize the result of your research into each question.

Please note - you are *not* being asked to *answer* the question posed in each situation. You are asked only to decide what kind historical explanation and truth is likely to result from your efforts to answer the questions. Mark either an A, B, or C in front of each question.

Feel free to use any knowledge you have about the question as you decide which letter to choose. You do not need to respond to the question if you have not had any contact with the historical situation in your AP US History course this year.

One final note: This is not in any way a test. Do not assume there is a preferred pattern of responses for you to select. You might find yourself choosing two of the letters more than a third one or you might find you've selected the same letter for all nine questions. The best approach is not to worry about responding in any "right" way and instead to focus on each question individually without thinking about what you marked for any other question.

Thank you for your help ... and have fun!

Please remember to complete the identification questions at the end

1) Andrew Johnson assumed the presidency under difficult conditions - a terrible civil war had just come to an end and the country had suffered the assassination of a sitting president. Though born in the South, he had supported the Union cause during the war. His poor background left him suspicious of both well-to-do whites and newly-freed Negroes. He engaged in a running battle with Congress for much of his time in office. Many felt he would be the first president to be removed through the process of impeachment.

_____ *To what extent were President Andrew Johnson's troubles due to his own actions versus the political climate of the times?*

2) The turn of the 20th century was marked by an expansion of industrial activity in the United States like none before it. This period - roughly the last twenty-five years of the 1800's and the years before World War I - saw the creation of manufacturing and financial empires by a number of individuals of great entrepreneurial ability. Two of the best known of these powerful men were Andrew Carnegie and John D. Rockefeller.

_____ *To what extent were the successes of these two early businessmen due to their managerial genius as opposed to their being in the right place (the United States) at the right time?*

3) In the history of the presidency, we justifiably give the most attention to those who have actually become president. There is, however, a continuing interest in candidates who lost presidential elections. Some presidential losers appeared to have the election won before the voting (think of Thomas Dewey in 1948). Others never seemed to have had a chance of winning. The reasons for winning - or losing - a presidential election are many. Consider two elections in the early part of the 20th century. Both Al Smith (1928) and Alf Landon (1936) were defeated by landslides in both the popular and electoral vote.

_____ *What led voters to choose the other candidate by such a wide margin in these two presidential races?*

4) The idea that there was an almost divine force at work to expand the sovereignty of American democracy was seen in many events as the 19th century neared its mid-point. The Mormons were moving west. A treaty with Great Britain resolving the Oregon Controversy provided a northwest anchor in what is now Washington state. In May, 1846, following a skirmish on the disputed border of southern Texas, President Polk asked Congress to declare war on Mexico. Less than two years later, the United States and Mexico signed a treaty that added some 500,000 square miles of former Mexican territory to its southern and western borders.

_____ *If we think about North America as a whole, to what extent did the actions of the US government at mid-nineteenth century contribute to political and economic stability?*

5) The 1920's were a time of many changes in American life. Wary of old-world culture and rivalries, our government promised a return to the "normalcy" of international isolationism. At home, there was growing prosperity - at least for those at the top of the economic ladder. It was the time of both prohibition and flappers. It has been suggested that modernization in the modes of economic production, along with tensions between urban and rural economic interests, together explain much of what was different about the 20's from the years before.

_____ *To what extent did changes in economic factors account for the changes in social structure and social relations in the US in the 1920's ?*

6) In 1786, the United States of America was more a collection of nearly-sovereign states than federal union. Many citizen-soldiers of the Revolution had not received their pay, and creditors were demanding payment for past loans. In Massachusetts, the situation resulted in six weeks of fighting in January of 1787 between farmers and the troops of Governor Benjamin Lincoln. The fighting became known as Shay's Rebellion, named for former colonial Captain Daniel Shays who became the chief spokesman for the farmers.

_____ *What happened from the end of the revolutionary struggle with England until 1786 to make former patriots turn against their new government?*

7) The presidential election of 1968 occurred during a time of national turmoil. The politics of the year reflected the events and mood of the nation. President Johnson announced his decision not to run for re-election in March. Governor George Wallace ran an independent campaign for president. There was rioting in August at the Democratic convention in Chicago. In November, Richard Nixon was elected president by a razor-thin plurality.

_____ *Was Nixon's election in 1968 a victory determined by traditional party loyalties, or was it the result of a backlash against the liberal cultural values presumed to be supported by the Democratic party?*

8) World War II was not yet over when a meeting to define the charter of the United Nations took place in San Francisco in the spring of 1945. Yet despite this early sign of cooperation, the period after the war suffered from the legacy of pre-war rivalries among the world's nations. In March, 1946, Prime Minister Churchill declared an "iron curtain" constructed by the Soviet Union had descended across eastern Europe. For nearly fifty years, a condition of mistrust and suspicion known as the Cold War existed between the United States and the USSR.

_____ *Compared to what we know about the acts of the Soviet Union, what can we say about the role of the US in starting and sustaining the Cold War?*

9) By the end of the Civil War, much of the land in the South was unfit for farming and manufacturing capability was almost non-existent. The Freedmen's Bureau, intended to assist the former slave population, operated for only one year. Simply stated, the economic and political process of reconstructing the nation was daunting. Amid bitter debates about how to go about accepting the breakaway states back into full membership in the Union, radical reformers from the north headed south in large numbers to become involved in politics at the state level. These so-called "carpetbaggers", joined by southern "scalawags" sympathetic to the northern Radicals, gained control of many southern state legislatures.

_____ *Overall, were the intentions of these politicians motivated out of self-interest or beliefs that their work would benefit the people of the south?*

Appendix B

Questions and Information

When a question is asked or there is a problem to be solved - both within and outside of school - people have different ways of seeking and organizing the information they will use to answer the question or solve the problem. The scenarios below were created as examples of the kind of situations a person like you might encounter in your life as a student today. Please try to respond to each question fully - that is, list five responses to each question.

Please complete the identification questions below before you begin !

Thank you for your help!

1) You saved all the money you have earned from your part-time job the last three years (you also won a cash prize in an essay-writing contest!). Now you are finally able to buy a car of your own. You begin the search by looking in the "used cars for sale" section of the newspaper. You see an ad for a car in your (limited!) price range. You call the number. What five questions would you ask the owner on the telephone that would give you a good idea of whether you actually wanted to see the car?

* each question is followed by five sets of two-lines for entering responses *

2) It is Monday morning. You are looking forward to lunch so you can talk with friends who had gone to a concert given by one of your favorite groups on Saturday night (you were late calling for tickets and the concert had already sold out). The line in the cafeteria moves slowly, so you will only have a few minutes to talk by the time you sit down. As you approach the green jello offered for dessert, you think of several questions about the concert to ask your friends. What are the first five ... ?

3) Imagine you were volunteering as a member of a search and rescue team in the White Mountains and you found two lost and injured hikers and helped carry them to safety. The members of the team have decided to get together after dinner to talk about the events of the day. As you get ready for the discussion (and plan to claim your favorite place on the couch by the fireplace), which of the day's events (and what about them) would you be thinking about? List five ...

4) You've been selected to be on a committee to make recommendations to the school board about whether uniforms should be required for all students in the schools of your district. Up to now, you haven't given much thought to this matter. But to your relief, another member of the committee gives you the numbers of several schools that had discussions about school uniforms last year. As you get ready to call these schools (using the school board's credit card!), what are five questions you would ask to help you prepare for the committee's first meeting?

5) In the presidential campaign this year, we heard the candidates use words and phrases such as "freedom" and "family values". In talking with your friends and classmates, you realize political discussion sometimes gets confusing because not everyone has the same sense of what such terms mean. Some people feel that such terms have (or should have) fixed meanings. Others believe these terms are best left open to interpretation depending on the intent of the person using them. Give five reasons to support one or some combinations of both positions regarding the need for relative or exact meanings for such political terms.

6) The spread of AIDS is a major social concern. You have been asked to help a planning committee develop an AIDS-awareness program in your town and need to deliver a short speech to the committee tomorrow night. Think about what you would advise the committee to emphasize in the program ... what topics or themes come to mind? Please list five.

7) As a leader of a teen-counseling group, you have been asked to mediate an on-going conflict between two sixth-graders in your neighborhood. Each child claims the other was at fault in the beginning (the dispute has been going on for several weeks). As you think about how to begin the mediation process, what are the first things you would do? Please list five.

8) Imagine you are a counselor at a summer camp. Your campers are middle-school age. You have been asked to explain and demonstrate a craft or a sports-skill at a special all-camp meeting. What craft or skill would you choose? As you think about how to begin, what are the first five things you would say or do in the way of explaining and demonstrating?

the craft/skill: _____

9) A club (or a team or a church group) you are a member of has decided to use the money raised from last year's candy sale to take a trip to New York City over a long weekend. A travel agent will organize the trip - arrange for travel, hotel and meal accommodations, and activities - to meet the wishes of the group. You have been asked to make contact with the travel agent and report back to the group before making final plans for the trip. What are five questions you would ask the travel agent?

10) A market research company has brought together a group of people to discuss how they reach conclusions about which brands to buy. The first question for members of the group concerns how consumers think about what makes a particular brand a "good deal" and worth buying. Specifically, you are asked to think about something you might like to purchase. What is it? Approximately how much does it cost? What questions could you ask in order to know which specific brand of this item was a "good deal" if you were to purchase it. List five questions ...

Item _____ Cost _____

Appendix C

Narrative and Paradigmatic Modes of Thought

A Coding Guide

The purpose of this guide is to provide the information necessary to develop the ability to place responses to the ten open-ended questions that make up the *Structure of Knowledge* questionnaire into one of two categories - the categories of narrative and paradigmatic modes of thought. The psychologist Jerome Bruner has claimed the categories represent two “irreducible modes of cognitive functioning”, two natural-kind ways of organizing perceptual input.

Narrative and Paradigmatic Modes of Thought

Bruner clearly states his central thesis: *there are two ways of organizing the sensory data we take in from the natural and social world around us*. The two ways aren't necessarily in opposition to each other, and one is not in any way better than the other. They each help us to make sense of the world around us and to describe that world to others. In one sense, the two modes of thinking operate to create worlds that exist in parallel dimensions - that is, if one person was thinking in one mode and a second person was using the other, even though they were both trying to characterize the same set of events, since they were organizing aspects of the situation differently, the “reality” they were describing would be different. Bruner believes these two independent ways of thinking should not be thought of as contradicting each other - they are simply different ways of organizing information - of applying a different set of beliefs about the way the world works to the whole range of sensory information we take in from the outside world.

Each mode of thought - each way of organizing the world - has its own set of criteria to describe how it is carried out, and each has its own standards for indicating the goodness or truth of the results of such thinking. It is these criteria and standards that will be used as guides for categorizing responses to the questions on the *Structure of Knowledge* questionnaire.

The Paradigmatic Mode

One manner of thinking is the *paradigmatic*. Bruner refers to this as the logico-scientific mode. The paradigmatic mode of cognitive functioning seeks to make meaning - to explain the world - by means of timeless and universally applicable statements of proof. Paradigmatic thinking is about taking in information and sensing what is common among pieces of data, thus organizing the information into categories.

As we develop these cognitive categories, we may also identify relations between and among the categories themselves. For example, the categories of “dogs” and “cats” and “cattle” are similar in that they all contain animals that have four legs and are warm-blooded. The first two categories (“dogs” and “cats”, but not “cattle”) can be included within a wider category which could be called “domestic pets”. Such first and second level category construction can lead to elaborate representation of worlds by the use of symbolic logic as a connector between and among individual categories.

Paradigmatic thinking is often seen in the domains of mathematics and the sciences. When done well, it leads to good theory and proofs guided by the application of formal logic. Good paradigmatic thinking can also lead to the discovery of new information (think of how theories in chemistry and physics guide the discovery of new chemical elements). Thinking paradigmatically means thinking inductively in a building-block way, seeking principles of generality.

The Narrative Mode

The *narrative* mode of thought focuses on the desires and intentions of people (or objects that are treated as if they were people ...), their actions, and the consequences of their actions. Narrative thinking involves thinking about the specific setting within which human action takes place - the physical and social context for action. It is also concerned with knowing how those in the action think and feel. Bruner believes that fully formed narrative thought cannot exist until a person can both list the events of the action and understand the intent of the actors.

Organizing thought in a narrative way explains the world by providing a way to know what those in it do and believe. In fact, the mark of good narrative thought is that it is believable - believable not necessarily in the sense that narratively organized thinking corresponds to some external reality, but believable because it connects and corresponds with our beliefs about why people act the way they do (or, at least, we *wish* they would act). Good narrative thinking makes sense in a “lifelike” way that is seen in good storytelling.

The narrative mode often has to do with issues of good and bad - of values. Yet, despite the universal quality of such normative matters, the building-blocks of narrative

thought are feelings and intentions which gain their meanings only as they exist in a unique situation or setting.

Bruner claims each type of thinking occurs naturally without forethought, that each is easily recognized and, furthermore, that each type can be analyzed into constituent elements which make up that particular mode. These parts of the overall process are important for the purpose at hand - they will serve as the markers for identifying which type of thinking is represented in the responses we are coding.

The Components of Narrative and Paradigmatic Thinking

Bruner provides a rich (and not exactly easy to follow!) description of the different dimensions of narrative and paradigmatic thinking. Some of his constituent categories are at the same level of meaning as each other. Others are nested within overarching conceptual units. We can put this definitional confusion aside, however, and concentrate on three related, though slightly different ways to unpack Bruner's ideas. The first is to look at two major categories used to organize elements of thinking. The second is a brief statement regarding the goal of both modes of thinking. And the third is a check-list of characteristics which can be used as guides to distinguish between narrative and paradigmatic thinking. In each case - in the categories, the goal statement, and throughout the check-list - Bruner expands on the core idea that the events of the world can be interpreted in one of two ways - as either the working out of human intentions, or the developing of a structure of abstract causes for an ordered world that exists independent of human intervention.

The Organizational Categories

Language Selection

What Bruner calls the "vertical" form of linguistic organization - the choice of a word or phrase with the aim of maximizing either the denotative or connotative sense - is the first dimension that can be used to identify a response as either narrative or paradigmatic.

Narrative language has what Bruner calls "metaphoric richness" and is intended to evoke images beyond the literal meaning of the words themselves. Narrative thought intends to mean more than is said - to create a text open to interpretation. The meaning for words used narratively come from how well they relate to the sense of the words around them.

Words used paradigmatically have what Bruner calls a "singular and definite" sense. The intent of words used paradigmatically is to evoke a sense of meaning in the person receiving the message that precisely and exactly corresponds to the sense intended by the sender. Paradigmatic thoughts are as unambiguous as possible. People thinking paradigmatically attempt to mean what they say. In the paradigmatic mode, words used gain their meaning by how well they correspond with other words which themselves are

understood in the same way by the speaker and listener, the writer and reader ... the sender and the receiver.

Language Combination

Larger collections of information - complete responses to a survey question or directions to a distant destination or a story about a young woman leaving home - can be characterized by the manner of interaction among three categories used by Bruner to describe the basis for the formation of concepts. These conceptual building blocks are the *factive*, the *formal*, and the *functional*. Factive is used in the sense of affective; that is, it is used to indicate reactions of people. Formal is used to denote the intent to indicate order and organization. Functional refers to the unfiltered and uninterpreted goings on in the real world.

For Bruner, narrative thinking concentrates on the relationship between what's happening in the world (the functional) and how people respond to it (the factive). Paradigmatic thinking has its origins in functional activity. But paradigmatic thought is more than paying attention to the world. It is functional activity that is informed by the more abstract order of the formal.

The Major Question

Bruner believes that both narrative and paradigmatic thinking have the explaining of *causation* as their ultimate aim. But the types of causality are different in the two modes. In narrative thought, the intent is to explain the particular connections between two specific events. The paradigmatic search for causality is the search for general/universal causes. Good narrative thinking (good "story telling") is marked by telling how events can have a psychically-intended cause. Good paradigmatic thinking (good "hypothesis testing") results in testable hypotheses. Narrative thinking seeks the meaning of experience. Paradigmatic thinking seeks universal truth.

A Quick Check List

Bruner doesn't provide a way to distinguish narrative from paradigmatic thinking based on particular characteristics common to both modes and expressed differently in each. Nevertheless, without any category labels, the following is a list of what might be called characteristics (or components ... or descriptors) of thinking as they are expressed in each mode. Much of what follows has been presented above - the list serves as a good check-list tounderstanding of the differences between the two modes of organizing information.

Narrative thinking:

is connotative - the meaning exceeds the simple definition of the words.

is concerned with the reaction of people to events.

is focused on human (or human-like) intent.

defines meaning based on context - how the word is used in the particular situation.

is temporal - the passage of time contributes to meaning.

uses the elements of drama - feelings, intentions, and situations.

incorporates values into explication.

Paradigmatic thinking:

is denotative - the intended meaning is found in the exact definition of the words.

is concerned with the abstract ordering of the events themselves.

is focused on outcomes brought about by a logical chain of events.

establishes meaning based on how the idea corresponds to the meaning of external ideas.

is timeless - time is not a factor in determining truth.

uses facts and the relations among them as explanatory building blocks.

is value-free ... logic prevails.

Examples of Narrative and Paradigmatic Responses

Having read a description of Bruner's conception of narrative and paradigmatic modes of thinking and seen how different characteristics of thinking can be expressed in each mode (the quick check-list), you can begin to code responses to questions which are intended to elicit examples of either narrative or paradigmatic thinking.

Perhaps the best way to distinguish between narrative and paradigmatic responses is to look to whom or what the response is referring. Narrative responses will almost always contain reference to a *person*, to what he or she is doing or thinking about. Narrative thinking involves the particulars of these situations of human interaction. Paradigmatic responses refer to descriptions of objects and conditions in the natural world - such responses are the building blocks of more abstract descriptions of the world.

The language of narration is expansive - it is connotative, it is language that is open to interpretation. Narrative thought involves feelings and desires and values. It seeks to explain by referring to the intentions of people involved in human interaction. On the other hand, the language of paradigmatic thought is more precise. It is language that is intended to convey meaning in an exact way. Paradigmatic thought is the stuff of facts - statements that describe objective states of affairs. The paradigmatic mode also seeks to explain, but it does so by showing how physical or logical connections can effect changes independent of human involvement.

Though narrative thinking is people-centered, it is not sufficient to say that any statement having to do with people is narrative in form. A simple physical description of a person can be considered paradigmatic. Narrative thought is identified by reference to intent - it starts when we ask what the other person is thinking and why.

Narrative Responses

EXAMPLE from question #8 (the summer camp question):

"I would allow them to be creative and make their own design."

This narrative response shows the elements of human intent, interaction among people, and open-ended problem solving.

EXAMPLE from question #7 (the mediation question):

“To mediate I would hear all of the problems and then think of ways to solve them”

Again, this is a narrative response. It involves human activity (hearing and thinking ...). It refers to human-mediated problem-solving in a particular setting.

EXAMPLE from question #10 (the market research question):

“Are the brand of jeans in style?”

The student picked a pair of jeans as the item s/he wanted to purchase. Because this is a straight forward response (a yes/no response) it may appear that it would be the result of paradigmatic thought. Yet, the question asks about a condition that is the result of human activity (the creation of “style”) and is therefore considered as an example of narrative thinking.

EXAMPLE from question #8 (the summer camp question):

“Ask questions regarding what I had said and see if any clarification was needed.”

This is a summary question (it was the last response given) and reflects the student’s interest in attending to the interests of other people (in this case, the campers s/he was teaching). It picks up on the possible uncertain nature of human interaction (was the student clear in what they intended to say?) and is therefore categorized as narrative in nature.

EXAMPLE from question # 7 (the mediation question):

“ Tell each of the sixth graders that you will try not to become too much of a middle man - you will just keep things calm and to the point.”

This response is coded as narrative because it involves the feelings and thoughts of people. Note also how it refers to a specific situation - it is concerned with one setting, not an abstract and conceptual structure.

Paradigmatic Responses

EXAMPLE from question #10 (the market-research question):

“Is it sold in many stores?”

This response could have been given for almost any item chosen (the specific example was a pair of jeans). It was coded as paradigmatic, not just because of its form (a yes/no question), but because the answer to the question would provide information about a non-human condition and would help the person asking the question presumably make a logical decision about whether or not to purchase a specific brand of a product.

EXAMPLE from question #7 (the mediation question):

“Tell the kids - NO FIGHTING or NAME CALLING.”

This is a paradigmatic response. Though it involves human action (directions to guide behavior), it is clear and exact in meaning and is presumably part of a framework for action that is independent of any particular situation or setting.

EXAMPLE from question #8 (the summer camp question):

“Describe the necessary tools”

The student who gave this response was telling the campers how to bake a pie. The response referred to the physical world. It was the starting point of an explication that could be developed independent of personal and interpersonal thoughts and feelings. The reference to the tools was presumably intended to be unambiguous. It is a paradigmatic response.

EXAMPLE from question #10 (the market-research question):

“How loud does it go?”

The respondent who gave this answer was thinking about purchasing a CD player. The response is paradigmatic. Though the answer to the question is more open-ended than categorical (the CD player isn't just loud or soft), it still refers to a characteristic of the physical world. An answer to the question would likely refer to a scale that builds on physical phenomena and uses categories of description abstracted from these phenomena (e.g., the use of decibel notation to categorize the loudness of sound).

EXAMPLE from question #8 (the summer camp question):

“First, I would show a picture of the field and explain the names of (the) lines.”

This paradigmatic response is notable for the manner in which it sets the stage for an abstract representation of an idea - in this case, the structure of and rules for playing the game of soccer. There is no reference to human intent. The idea of the shape and description of the field exists “out there” in a realm of rules that are independent of any individual’s idea of how the game should be played.

Uncodable Responses

Very few responses will be uncodable (in fact, because the sample responses you will code were chosen at random, the set of 120 responses may not contain *any* uncodable responses). Responses that are uncodable will most likely be of two types - *unintelligible* and *frivolous*.

Unintelligible responses simply won’t make enough sense for you to categorize them as either narrative or paradigmatic. Words may be so mis-spelled that you are unable to make a reasoned guess at the intended meaning. Syntax may be so poor that you are unable to determine the sense of the phrase or sentence.

It is also possible that a respondent may simply not care to produce a thoughtful response. The instructions on the instrument ask them to give five responses to each question, and it may be that they have lost interest by the time they enter a response. In this case, the words and structure of the text may make sense, but it may not refer to the question posed, or be satirical or otherwise critical of the question or the instrument.

EXAMPLE from question #4 (the school uniform question):

“What were some of the reasons for people for(or perhaps “in”) under”.

This response is not intelligible to a reader - it has no sense or meaning. Hence it would be coded as U.

EXAMPLE from question # 2 (the concert question):

“Were they wearing green?”

The meaning of the words are clear, as is the syntax of the statement (recall the reference to green jello in the question!). But the response was coded as a U because it was decided that the student had made the statement simply to satisfy the requirement that s/he enter five responses to the question.

EXAMPLE from question #3 (the rescue question):

“What’s for dinner?”

Certainly an understandable statement (and cute!), but one that doesn’t respond to the question and hence coded as U.

Appendix D

The Bem Sex-Role Inventory

Directions

On the next page, you will find listed a number of personality characteristics. We would like you to use those characteristics to describe yourself, that is, we would like you to indicate, on a scale from 1 to 7, how true of you each of these characteristics is. Please do not leave any characteristic unmarked.

Example: sly

Write a 1 if it is **never or almost never true** that you are sly.

Write a 2 if it is **usually not true** that you are sly.

Write a 3 if it is **sometimes but infrequently true** that you are sly.
Write a 4 if it is **occasionally true** that you are sly.

Write a 5 if it is **often true** that you are sly.

Write a 6 if it is **usually true** that you are sly.

Write a 7 if it is **always or almost always true** that you are sly.

Thus, if you feel it is sometimes **but infrequently true** that you are "sly", never or **almost never true** that you are "malicious," always or **almost always true** that you are "irresponsible," and often true that you are "carefree," then you would rate these characteristics as follows:

<u>Sly</u>	<u>3</u>	<u>Irresponsible</u>	<u>7</u>
<u>Malicious</u>	<u>1</u>	<u>Carefree</u>	<u>5</u>

Sample items from the Short Form of the Bem Sex-Role Inventory (Copyright 1978, 1994 Consulting Psychologists Press. Used with permission from Mind Garden, Inc., Palo Alto, CA).

Short Form

1	2	3	4	5	6	7
Never or almost never true	Usually not true	Sometimes but infrequently true	Occasionally true	Often true	Usually true	Always or almost always true

Sympathetic (coded as feminine)

Understanding (coded as feminine)

Willing to take a stand (coded as masculine)

Independent (coded as masculine)

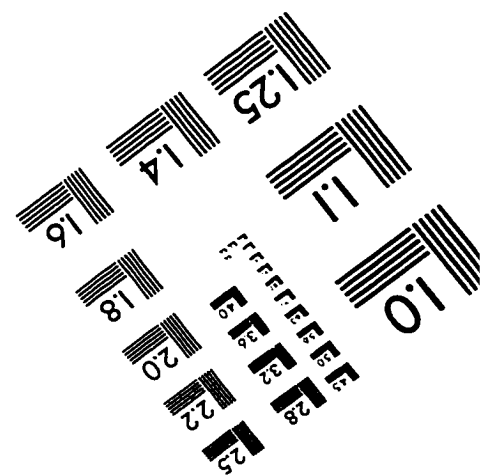
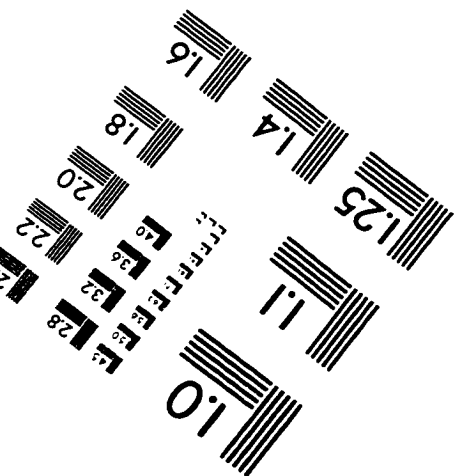
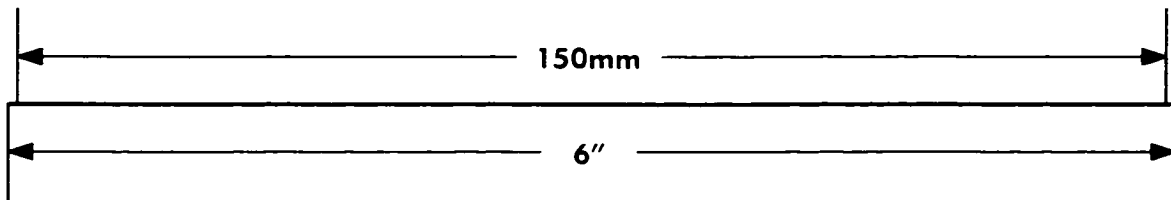
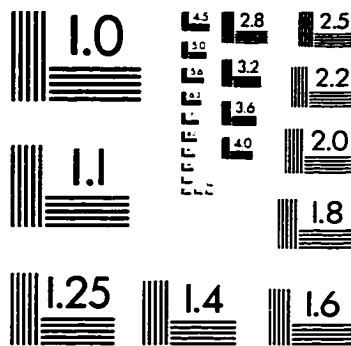
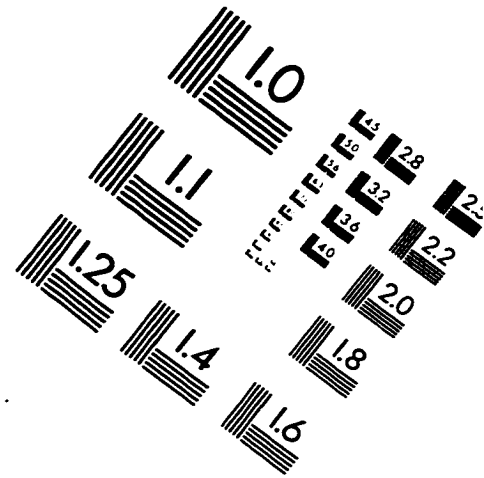
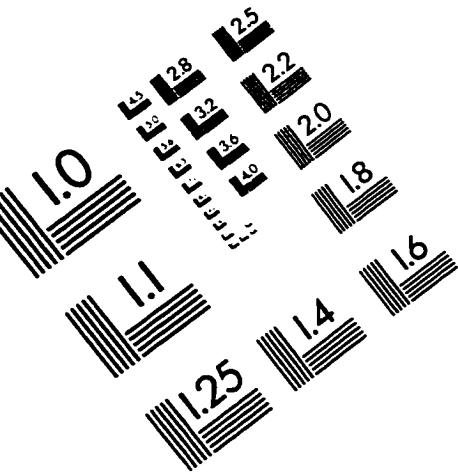
Adaptable (filler item)

Appendix F

The material included in Appendix F -- Section II, Part A (the Document-Based Question) and Section II, Parts B and C (the Essay Questions) from the 1997 AP US History Examination -- are not reproduced here for copy by University Microfilms International per request of the Educational Testing Service.

In other copies of the dissertation, Appendix F is displayed on pages 168 and 169.

IMAGE EVALUATION TEST TARGET (QA-3)



APPLIED IMAGE, Inc
1653 East Main Street
Rochester, NY 14609 USA
Phone: 716/482-0300
Fax: 716/288-5989

© 1993, Applied Image, Inc., All Rights Reserved