COGNITIVELY SENSITIVE ASSESSMENTS
OF STUDENT WRITING IN CONTENT AREAS

Eva L. Baker, Pamela R. Aschbacher,
David Niemi, Edynn Yamaguchi, Yujing Ni

Training package developed by the National Center for
Research on Evaluation, Standards, and Student Testing
UCLA Graduate School of Education
145 Moore Hall
405 Hilgard Ave.
Los Angeles, CA 90024-1522
(213) 825-4711

May, 1991
# TABLE OF CONTENTS

Acknowledgement  
Part I: Introduction  
Part II: Cognitive Assessment of Subject Matter  
Part III: Guidelines for Using CRESST's Cognitively Sensitive Assessment System  
Part IV: Sample Assessment Materials  
1. Teacher's Packet  
2. Student's Packet: American History/Civil War  
3. Student's Packet: Chemistry/Soda Task  
Part V: Specifications for Developing Assessment Materials  
Part VI: Rater Training, Scoring, and Reporting Procedures  
Part VII: Sample Training materials  
1. Prior Knowledge Scoring Materials  
2. Essay Scoring Guidelines and Record Forms  
3. Student Essays and Scoring Keys
ACKNOWLEDGEMENT

We wish to thank the many people who participated in the research and development activities that provided the basis for this handbook. Robert Linn, Co-Director of CRESST and professor of education at the University of Colorado provided invaluable technical guidance to the project. Marie Freeman and Serena Clayton contributed significantly to earlier stages of the project. Other UCLA colleagues and graduate students, including Regie Stites, Simon Chang, Jamal Abedi, and Howard Herl devoted much time and effort in field testing and data analysis.

Our sincere appreciation goes also to the students, teachers and administrators in the many schools in California, Illinois, Michigan, and Ohio where our materials were field tested, and to the many teachers who scored hundreds of essays for us. Without their willingness to participate, this handbook would not be possible. We also appreciate Ed Roeber and Tom Kerins, of the Michigan and Illinois State Departments of Education, respectively, for facilitating our data collection efforts in their states. And finally, we thank Lynn Winters, Director of Research and Evaluation of the Palos Verdes School District, for her very helpful review of an early draft of this handbook.
PART I
INTRODUCTION
PART I: INTRODUCTION

This handbook presents an alternative approach to assessing students' deep understanding of subject matter content. It is based on years of research conducted by the National Center for Research, Evaluation, and Student Testing (CRESST), funded by the U.S. Department of Education's Office of Educational Research and Improvement (OERI). Our approach assesses the content quality of student writing in response to primary source materials, and it utilizes a method of scoring that was derived from analyses of subject matter experts' and novices' writing.

This method was designed to be useful for two purposes: (1) large scale accountability, and (2) the diagnostic improvement of instruction. Typically, measurement experts have argued that accountability and diagnosis should be conducted with separate kinds of assessments. But for practical, economic, and conceptual reasons, we argue that accountability and diagnosis can be merged in a single measure, with different methods of reporting the data for different purposes.

The handbook provides training in our alternative assessment method and examples of our materials for your use. We recognize that our tasks, selected as they were to represent critical areas in American history and to meet the convenience and classroom contexts of our tryout samples, may not explicitly meet your needs. To that end, we have included specifications in Part V, designed to provide a template for the generation of tasks such as those we used. We would expect that the scoring scheme we developed would apply equally well to locally developed tasks that followed our model.

The examples presented in this handbook were drawn primarily from our work with eighth through twelfth graders on a topic about the Civil War era in U.S. history, along with some examples of student materials on a chemistry topic, analysis of an unknown substance. CRESST has developed complete sets of materials for other U.S. history topics (two on the American Revolution and one on the Depression) and chemistry topics, and we are beginning to extend this approach to other sciences, math, and geography as well.

Interested users may contact CRESST (213/825-4711) for copies of additional materials, assistance using them in an assessment program, help in developing new topics, or related technical information about the scales.

This handbook includes background information about the conceptual framework for the project and its development over the past several years, directions and sample materials for collecting data, and directions and sample materials for training raters, conducting the scoring process, and reporting results. This information is organized into the following sections:

Part II: Cognitive Assessment of Subject Matter: Understanding the Marriage of Psychological Theory and Educational Policy in Achievement Testing
Part III: Guidelines for Using CRESST's Cognitively Sensitive Assessment System
Part IV: Sample Data Collection Materials for Students and Teachers
Part V: Specifications for Developing Assessment Materials
Part VI: Rater Training, Scoring, and Reporting Procedures
Part VII: Sample Training Materials (including scoring guides, sample score recording forms, scored and annotated student essays and unscored versions).
PART II

COGNITIVE ASSESSMENT OF SUBJECT MATTER:
UNDERSTANDING THE MARRIAGE OF PSYCHOLOGICAL
THEORY AND EDUCATIONAL POLICY IN ACHIEVEMENT TESTING
PART II: COGNITIVE ASSESSMENT OF SUBJECT MATTER:
UNDERSTANDING THE MARRIAGE OF PSYCHOLOGICAL
THEORY AND EDUCATIONAL POLICY IN ACHIEVEMENT TESTING

Eva L. Baker, Marie Freeman, and Serena Clayton

UCLA Center for Research on Evaluation, Standards, and Student Testing (CRESST)

This chapter will describe the policy context of standardized testing and then detail theoretical precepts and an extended example of a new measurement approach.

Introduction

For at least a quarter of a century, educators and critics have raised conceptual and technical questions about standardized achievement tests (Strenio, 1981). And for the most part, the public and its policy makers have ignored these ululations and continued to believe in the accuracy and usefulness of these measures, dismissing technical concerns as abstrusely academic and teacher complaints, at minimum, as self-serving. However, recent reform efforts, stemming from A Nation At Risk (National Commission for Educational Excellence, 1983) and other dark reports of American educational quality, have directed renewed attention and investment in achievement outcomes. With the statement of national educational goals by the President in 1990 and the governors of the fifty states in 1989, and the President's promise to measure achievement in Grades 4, 8, and 12, standardized achievement tests are about to become national educational policy. The consequences of error in test design and interpretation are inestimably higher than in the past, for such measures will exert dramatic control on the public school curriculum, on what tests are published, and on what is taught. Information from achievement measures must answer three questions: What is the quality of our students' achievement? How can achievement be improved? Why can't present tests do the job? For the purposes of accountability and instructional improvement, the vast majority of existing standardized achievement tests are wholly inadequate. They create the wrong expectations and incite inaccurate inferences in terms of policy action. They are inappropriate in at least three central ways: their underlying theory, their content, and their procedures. These assertions deserve at least brief elaboration.

The measurement assumptions of standardized tests rely on models based upon theories of stable individual differences. These models posit a general construct such as mathematics ability or reading comprehension. Construct measurement has at least two requirements: (a) substantial variation among people on the target test in order to differentiate scores, that is, scores on the 78th or 64th percentile are intended to reflect different levels of performance; and (b) stability of measurement for individual performance for accurate prediction. When mapped against the requirements for assessing an individual's educational improvement or the impact of systemic educational reform, these instruments do not measure up. Reports from most standardized tests obscure the meaning of the test scores from the teacher, the student, and the public. We may know

---

1 The research reported herein was conducted with partial support from the U.S. Department of Education, Office of Educational Research and Improvement, pursuant to Grant No. G0086-003. However, the opinions expressed do not necessarily reflect the position or policy of this agency and no official endorsement by this agency should be inferred. A version of this document appears in M.C. Wittrock and E.L. Baker (Eds.), Testing and Cognition. Englewood Cliffs, NJ: Prentice Hall, 1991.

2 We wish to thank Tom Kerins, John Craig, and Carmen Chapman, of the Illinois State Board of Education; Bob Hill, of the Springfield Public Schools; Lynn Winters, of the Palos Verdes School District; and the many principals and English and history teachers who participated in or helped with this project. In addition, we would like to thank colleagues at UCLA who helped with various aspects of the study: Pam Aschbacher, Jamal Abedi, Joan Herman, Edys Quellmalz, Merl Wittrock, Simon Chang, Yujung Ni, Reggie Stutes, Kim Kyung-Sung, and Rebecca Frazier.
requirements for assessing an individual's educational improvement or the impact of systemic educational reform, these instruments do not measure up. Reports from most standardized tests obscure the meaning of the test scores from the teacher, the student, and the public. We may know the relative position of individuals and school districts compared to other individuals or school districts, but we do not know what level of performance any given score describes. Further, even under the best conditions, educational reform has weak effects. So to detect change, progress toward national goals, for example, achievement measures must be created that are sensitive to minor, but real differences in performance. Tests should tell us who has changed in ability to perform particular tasks at described levels of expertise. Standardized achievement tests do not tell us what we should want to know.

The problem of interpreting these tests is amplified by the way their content is selected. A major problem is content sampling within a particular subject matter, such as history or mathematics. Most subject matter measures are commercially available and are intended to be sold to school districts and states. To be competitive, testing companies must attempt to include a sufficient number of topics with broad appeal in any subject matter. A common result, as predicted two decades ago (Popham & Husek, 1969), is a content-curriculum mismatch, where the overlap on test content and curriculum varies by district, school, or classroom. This phenomenon has been documented in many specific subject fields, for example, in mathematics by Flden and his colleagues (1980). In practical terms, a mismatch means that certain topics that are untreated in the curriculum of given classrooms and schools will be included on the test. On the other hand, even topics emphasized in teaching may only be superficially measured because of time constraints.

Both types of errors result in misrepresenting students' actual achievement. One solution to this problem has been to encourage teachers to adapt their instruction to match test content (a process called "alignment"), a course of action that cedes enormous and inappropriate power to the developers of such tests.

A second, more global content issue is created by the pressure to test in a relatively limited number of subject matters. Such choices have been made as a matter of course to save money and time as well as to constrain the number of measures on which public accountability will be based. As a rule, districts and states commonly select an essential core of subject matter, often the areas of reading and computational skills in mathematics. Teachers and school policy makers adapt instructional time to focus on the goals to be measured. One consequence of this adaptation may be a reduction in time for untested subject fields: foreign language, the humanities, the arts, and the sciences. This reduction occurs logically to focus resources on the accountable aspects of the curriculum, but also because of the widespread, pernicious belief that students must learn the "basics" before they can profit from exposure to other subject matters and more complex intellectual processes. Particularly for poor performing students, opportunities in a wide range of subject matter are foregone (Oakes, 1986). The result is obvious—an educational system with clearly constricted curricula, differentially skewed to limit the access of the already disadvantaged.

The constraints on administration and scoring of standardized tests also influence their impact on school learning. Tests have been developed with strict time boundaries, partly in an effort to reduce testing time, partly from historical, psychometric reasons as result of their original purposes to differentiate among individuals. To obtain differentiated and reliable responses, it is better to limit test time and to expose students to many short items. More test items also mean more topics can be covered. Multiple-choice items are the most frequently preferred achievement testing format because they are time sensitive, permit responses to a relatively large set of items, have acceptable psychometric properties, and allow economic scoring approaches. How do these choices affect students? Multiple-choice formats exert undeniable control on school practice. The format frames how information is presented, learned, and retained. These tests assess learning in an artificial, decontextualized manner that is remote from how students learn or will apply knowledge in the future. These tests are likely to reduce student motivation to perform and are
likely to inhibit transfer. Such formats also convey a false sense of objectivity and quantification of performance, and objectivity and quantification are high-priority attributes in our society.

If it is true that such tests measure content only partially represented in school instruction and use formats convenient for administration and scoring rather than for learning, the simple problem of showing "improvement" in achievement is difficult and daunting. If it is difficult to design programs whose effects are detected by accountability measures, educators have few acceptable options. They may persist in doing the best they can, but may continue to see public confidence erode when test scores do not respond to their efforts. They may react in ethically questionable or unacceptable ways, for instance, selecting tests that seem easiest rather than those that most accurately measure valid educational goals (Cannell, 1987; Linn, Graue, & Sanders, 1989), encouraging inappropriate practice of items on the test (Shepard, 1989; Popham, 1990), or even falsifying test results. Schools may respond by offering training programs to develop test-taking skills and may confuse, once again, ends and means. How do students, the nominal object of our concern, respond? From the broad evidence, it appears with less interest and focus at best, and with active subversion at worst. Thus, with the best intentions, policy makers, compliantly supported by the public, require standardized achievement measures as the principal indicator of educational effectiveness and continue to deform the system in serious ways. Even a partial acceptance of this analysis raises serious questions about the quality of inferences we are drawing from standardized tests.

New Choice Points

The expectation that accountability measures will directly and productively influence student achievement is wildly optimistic. Their imposition influences broad instructional choices: how much time is allocated to various subject matters, and what particular topics are covered. But to affect important student performances, measures must influence a far deeper and dynamic level of instructional decision making. They must provide guidance and be sensitive to differences not only in what topics should be included in the curriculum, but also to the dirty details of teaching and learning, the instructional processes that differentially affect performance. Unless measures are ultimately sensitive to significant instructional choices, their impact on school improvement will continue to be marginal, periodically stunning policy makers who use terms like "stall" to explain the dysfunction between their own accountability fantasies and the actual utility provided by test results for day-to-day instructional planning. One answer has been to search for alternatives to the existing tests that will provide help to improve instruction. Unfortunately, this strategy has resulted in the propagation of tests functions with little linkage between them—for example, between diagnostic and accountability tests. We need measures that can provide information at the right level of detail to guide instruction but that will not divert large proportions of instructional time from learning tasks.

To meet the legitimate concerns for accountability and resulting instructional improvement, we require new approaches. It is time to break away from the inertia of present achievement testing practices, from the never-never land thinking that we can make schools better only by trying harder. We need outcome measures that simultaneously avoid the major deficits of standardized tests and provide trustworthy and useful achievement information. Critical attributes possessed by new cognitive approaches to testing are (a) their focus on important and teachable learning processes, (b) the confidence we can place in their measurements, and (c) the appropriateness of cues they provide for instruction.

Cognitively Sensitive Assessment

If we start with the notion that tests should measure significant learning in a way that supports desired performance, we are immediately led to a reversal of present practice. Instead of having tests constrain instruction, assessment procedures should be constructed to map directly on significant features of learning. Through close observation, skilled experts can tell whether learners are making progress on a wide range of intellectual tasks. Our problem is to transmute the
critical aspects of that observational process into procedures suitable for use in large-scale assessment. We must shift our view from the measurement of broad constructs to the assessment of important and described classes of cognitive learning tasks—knowledge acquisition, deep understanding, and problem solving. These processes must be assessed as they are embedded in various tasks and content domains; however, our assessment strategies may attempt to capture attributes of performance that transfer across subject matter domains. In our CREST project on assessing deep understanding of subject matter, we conducted research designed to transfer knowledge that developed in learning research and apply it to the problem of assessing the understanding of history. What will follow is a chronological description of the developmental history of our project, interpolated by discussions of the generalizable problems confronting developers of new approaches to assessment.

**Project Goals and Plans: Developing New Criteria for Scoring Writing in History**

Stimulated by articulate statements about the importance of knowledge of history by Hirsch, Kett, and Trefil (1987) and the dismal performance of American students on tests of historical knowledge (Ravitch & Finn, 1987), we decided to focus our attention on the measurement of history knowledge. We decided to attempt to assess a deeper understanding of history. We conceived of the problem for students as a comprehension task dependent upon their ability to generate or construct meaning (Wittrock, 1974) from provided stimuli by activating students' prior knowledge. This approach contrasts with the conception of history knowledge as a single construct dependent upon the accumulation of separate pieces of knowledge. Consequently, we broadened our approach from the usual multiple-choice format. Our research group had considerable experience in developing measures of writing skill (Baker, 1987; Quellmalz, Capell, & Chou, 1982). Our initial idea was to attempt to expand the content quality scoring rubrics used to assess writing and to apply them to subject matter topics in the field of history. Extant content quality scoring rubrics have treated content in one of two ways: as elaborated detail that contributes to a good essay in holistic scoring; or as important, unique material dependent upon the particular topic presented the learner. This second conception guides approaches used in scoring Advanced Placement Tests in History (Vaughan, 1983) and in primary trait scoring in the National Assessment of Educational Progress (1990). In this topic-dependent approach, individuals with expertise in the assigned topical area meet and develop post hoc standards for the particular set of papers written. The benefit of this procedure is the development of scoring scales that are particularly appropriate for the topic assigned. However, that strength is at once a severe limitation: First, the level of specificity required to adapt scoring criteria to a particular topic inhibits their more general use for other, similar topics. Thus, every topic possesses a unique set of criteria. Combining such particularized assessments across a range of topics or over a number of years involves a complex scaling process, based on equating results for different topics. Among a number of flaws, a major consequence of scaling is the ambiguity of score meaning. A second limitation relates to the inferences for instruction that can be derived from such measures. If every topic requires a unique set of criteria, what guidance can be provided to the teacher to inform teaching processes to improve student performance? Only if the tasks and scoring criteria are made public—released by the test producers—can teachers guide students to meet such standards, and then only if the same tasks are used. The trick is to find the appropriate level of generality to describe criteria so they are simultaneously appropriate for the particular assessment topics and conceived in terms that can guide future instructional practice and assessment.

**Goals**

The goals of our assessment research in the measurement of deep understanding of history were (a) to develop valid formats for eliciting students' thoughtful explanations about history concepts, (b) to create and validate content quality scoring criteria for students' responses, and (c) to explore these developments in the context of large-scale assessment settings. A longer term interest is to communicate the test design characteristics so that they will be helpful to the design of effective teaching strategies.
Strategies

Target. In light of our technical expertise in writing assessment, our project focused attention on essay writing in history. We believed that the strong tradition for this type of task in history instruction would increase the chances, if successful, of widespread acceptance of new assessment strategies. We also determined from reviews of plans for state assessment activities that writing in social studies was planned for many of the more forward-looking state assessment enterprises (for instance, California, Connecticut, Illinois, and Michigan). Finally, we believed that the present approaches used in the scoring of content-focused writing were inappropriate both conceptually and practically for the dual purposes of measuring deep understanding in large-scale settings and providing inferences useful for instruction.

Plan. In order to verify the need for essay scoring systems to assess content quality, we first had to determine if content specific scoring criteria for history already existed implicitly in the scoring behavior of history teachers. If so, we would identify these criteria, train others to use them, and validate their utility. If not, we would explore the literature to infer criteria that might be used. Even though our goal was to develop scoring approaches with reasonable generalizability across tasks to facilitate instructional improvement, we decided to limit our studies severely. We planned to focus on a grade level (11th grade) and on a single topic area in history, for we wished to be sure our findings were well grounded in a defined context. If we were encouraged by our results, we planned to test the generalizability of the approach: for other subject matter areas, for the age ranges of students for whom the approach was useful, and for sets of administration conditions. In sum, we anticipated the development of broadly useful assessment approaches as we conducted initial research in a restricted environment.

Our first problem was to identify specific content topics and strategy for data collection that would allow us to explore the issues of content quality scoring criteria. One requirement was to assure that students had some previous exposure to the concepts we planned to assess so that they could respond to our tasks. We hoped to assign passages in commonly used textbooks for this purpose. To that end, we reviewed textbooks, literature on the teaching of history, and available curriculum guides to determine the topics and most desirable sections of secondary school textbooks appropriate for our experiments in measurement. Our review of textbooks led to unoriginal but nonetheless disappointing results. For every topic we pursued, we discovered that secondary school texts presented relatively superficial treatments, without sufficient concepts and depth of supporting knowledge to allow the development of deep understanding. These views have been supported in the literature by Beck, McKeown, and Gromoll (1989), Sewall (1987), and FitzGerald (1979). We also consulted at length with the staff of the UCLA Center for the Study of Teaching and Learning in History, a collaborative enterprise of the National Endowment for the Humanities that brings together experts in history and curriculum.

Goal Redefinition

Because we were unable to identify suitable text segments for use in the assessment, we decided to incorporate the reading of a provided text as part of the assessment procedure itself. This decision transformed in a serious way our assessment focus. Rather than an exclusive focus on measuring the accumulation of information developed over a long period of instruction, we now attended to two major content issues: students' ability to read and integrate new information with previously learned knowledge, and students ability to explain new ideas using their prior knowledge. This transformation placed our work squarely in line with cognitive views of language comprehension (Anderson, Spiro, & Anderson, 1978; Rumelhart, 1980; Brown, Bransford, Ferrara, & Campione, 1983; Kieras, 1985). However, we were still driven principally by our subject matter concerns, a fact that guided the formulation of criteria for the topic and text selection for assessment tasks displayed in Table 1.
Table 1
Criteria for the Selection of History Texts to Assess

1. Must be a regular and significant piece of the secondary school history curriculum in the United States.
2. Must depend upon primary source material rather than summaries in textbook.
3. Must allow for multiple interpretations and inferences.
4. Must transcend immediate events and allow students to find relationships to other historical and contemporary events.
5. Must be brief enough to read within a class period.

Based on the application of these criteria, we decided that original speeches or essays composed by historical figures would meet criteria two, three, and five. For our initial set of studies, we selected the texts of the Lincoln and Douglas debates on popular sovereignty and slavery, choices that met the remaining criteria as well.

Identification of Content Quality Scoring Criteria: The First Pass

Our goal was to assemble valid criteria to assess understanding of history content. But essay writing consists of both content expertise and communication skills. We were well aware and troubled by the high intercorrelations in the literature between subscores on essays of expression skills and content knowledge (Baker & Quellmalz, 1980; Langer, 1984). Although it was obvious that highly verbal students would usually learn more about verbally based content areas, we were especially interested in discriminating performance between the ignorant facile writer with little subject matter understanding and the knowledgeable student with less developed writing skills. This desire corresponded to the common practice of high school teachers, who give both a "content" grade and a "form" grade (e.g., A-/B) on student essays. We wanted to focus on the elements that compose the content score.

A related concern was the impact of content knowledge (or lack thereof) on the raters' application of scores. We believed that knowledgeable people with experience in the subject matter would be needed to make the levels of distinction in which we were interested. Our first empirical study attempted to determine if the quality of content in essays, its accurateness, aptness, and structure, would be judged similarly by history teachers using implicit but common criteria for quality. We would contrast their ratings with those given by English teachers, specifically teachers trained to score essays in terms of the quality of general writing skill or expression, such as organization, style, and purpose. The essays we collected for this study were provided by 85 eleventh-grade Advance Placement (AP) history students in a suburban high school. We chose AP students because they would be likely to write "scorable" papers, that is, produce a sufficient quantity of writing to be graded. The AP students also had been exposed to an instructional sequence on the pre-Civil War period approximately five months earlier, so they would possess some background knowledge of the topic.

The experimental procedures spanned two consecutive days. On the first day students were given a general multiple-choice examination in pre-Civil War history, a test that had been validated by six expert history teachers. Next, students completed a background questionnaire describing their grades in English and social studies, self-estimates in ability, interest in writing and in social studies, and descriptions of teachers' instructional and assessment practices in history. On the second day, students were randomly assigned to read either the Lincoln or the
Douglas debate text. After the students completed their reading, they were given an essay question in either a brief or an extended form that asked them to explain the author's main issues and why they were important. Students were allowed 50 minutes to read the text of the speech and to write their essay. The papers were independently scored by two groups of raters: the English teachers and the history teachers.

**Procedures for English teacher raters.** One rater group was composed of four English instructors, all highly experienced in rating student essays according to holistic and analytic techniques. All had been trained to use the writing scoring scales developed at UCLA (Smith, 1978; Quellmalz, Smith, Winters, & Baker, 1980) and subsequently adapted for use in numerous state assessments, research studies, and the international comparisons of written composition performance (Baker, 1987). These scales included four major categories—general competence, essay organization, paragraph coherence, and support (meaning detail)—as well as scales for grammar and mechanics. We also were interested in the thought processes that raters used and their initial levels of stringency. Thus, we asked raters prior to their training to read three sample papers privately, to rate them on a five-point scale, and to comment on their decisions and impressions; comments were tape-recorded. Raters also were asked to identify criteria for a good paper. The training was conducted using procedures described by Quellmalz (1986) with model papers and illustrations of score points. The raters were told explicitly to focus on issues of presentation and rhetorical effectiveness rather than content-specific issues, such as content accuracy and depth of explanation. Nonetheless, during the training the raters insisted on modifying the scoring system. They decided to include as part the general competence subscore some index of the student’s attention to the specific writing task. All raters independently scored each of the 85 papers.

**Procedures for the history teachers.** Independently, and with no knowledge of the English teacher group or their resulting scores, a group of five history specialists was assembled to rate the same set of essays. Two were high school Advanced Placement teachers (from a school different than the data collection site) and three raters were advanced graduate students in history. Like the English teachers, all history raters were asked to assess three essays and to think aloud into the tape recorder as they completed this rating task. Their actual rating instructions differed dramatically from those given to the English teachers: No preexisting scoring scale was used, and no extensive training was conducted to determine if the history group shared implicit criteria. Each rater was told to give each paper two scores. The first score was to reflect how well the essay demonstrated serious understanding of the debate text read by the student. The second score was to provide an estimate of the essay’s general quality, taking into account issues other than the essay’s content. These scores conformed to the content-form scoring mentioned above. We also asked the history group to select the ten best and ten worst essays, so that we could infer from their choices the operational criteria they used to make their judgments. Each history teacher independently rated each of the 85 papers, giving each a content quality and an overall quality score. Following the rating session, all teachers discussed in a group the attributes that distinguished the highest from the lowest rated papers.
Findings and Interpretations

Detailed data analyses were conducted; only the highlights will be reported here. No significant differences on student performance were found for text passage (Lincoln or Douglas) or question type (brief or extended), in the ratings of either group. Our findings verified the inappropriateness of the existing UCLA scoring scale for the content focused task we used. Alpha coefficients among raters ranged from a low of .52 for mechanics to a high of .75 for general competence (the one score where raters took into account the task content). This finding reinforced the need for the development of a content quality scoring rubric. For the history raters, the alpha coefficient on general quality was .69 and on content quality was .75. The generalizability ratings for English raters (4 raters by 4 subscales) was .65 and for history raters (5 raters by 2 subscores) was .73. An interesting finding was that the percentage of exact agreement for scores given in the history group to content quality was only 33%, suggesting that no clear set of implicit criteria was operating among the history specialists. In addition, a t test was computed between average scores given by the history teachers and the history graduate students; significantly higher scores were assigned by the secondary school history teachers. The correlation between general competence scores assigned by English teachers and history content quality scores on the same papers was .80, similar to the relationship between general competence assigned by the English group and the general quality score assigned by the history group (.82). Such data suggested that English and history teachers were looking at papers in fundamentally similar ways.

Unfortunately, the expert knowledge possessed by history teachers did not seem to differentiate their judgments of student essays. But some aspect of special knowledge was operating, however faint. A low but significant correlation was obtained between the content quality scores of the history teachers and the total multiple-choice knowledge score (r = .32, p < .05). Leads for the development of content quality scoring criteria had to come from other sources. We then reviewed the history raters' think-aloud ratings and their post-rating discussions of the ten best and worst papers. The historians agreed that the best papers had the qualities listed in Table 2.

<table>
<thead>
<tr>
<th>Table 2</th>
<th>History Specialists' Generation of Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Established historical context</td>
<td></td>
</tr>
<tr>
<td>Presented a sound thesis early in the paper,</td>
<td></td>
</tr>
<tr>
<td>Detail contributed to thesis, was correct, and was not simply opinion</td>
<td></td>
</tr>
<tr>
<td>Avoided absolute judgments</td>
<td></td>
</tr>
<tr>
<td>Presented multiple points of view</td>
<td></td>
</tr>
<tr>
<td>Avoided interpreting the past in terms of present conditions</td>
<td></td>
</tr>
</tbody>
</table>

Scoring Criteria: Pass Two

In an effort to explore the utility of these criteria, a comprehensive and detailed scoring rubric was constructed based on these categories. The 12-category scoring scheme comprised the elements in Table 3 below; these elements were to be used as scoring dimensions for the papers.


<table>
<thead>
<tr>
<th>Identification of the Historical Problem/ Central Concept</th>
</tr>
</thead>
<tbody>
<tr>
<td>Depth of Elaboration</td>
</tr>
<tr>
<td>Breadth of Elaboration</td>
</tr>
<tr>
<td>Flexibility</td>
</tr>
<tr>
<td>Fluency/Detail</td>
</tr>
<tr>
<td>Evidence of an Analytical Problem</td>
</tr>
<tr>
<td>Goal Orientation</td>
</tr>
<tr>
<td>Logical Structure</td>
</tr>
<tr>
<td>Evidence of Historical Analysis</td>
</tr>
<tr>
<td>Autocriticism</td>
</tr>
<tr>
<td>Presentation</td>
</tr>
<tr>
<td>Style</td>
</tr>
</tbody>
</table>

Detailed descriptions for each of five scale points for every category were prepared. Based on a brief tryout with raters and reviews by experts, however, we deemed that this comprehensive set of categories was too ambitious. A review of literature on characteristics of expert knowledge (See Voss, 1978) suggested how we could pare the set down to five categories thought to represent critical attributes of historical thinking: Historical Context, Depth of Elaboration, Breadth of Elaboration, Evidence, and Historical Analysis. In addition, we added two categories related to expression, Rhetorical Structure and Mechanics, as well as an overall quality rating, General Impression. New scale point descriptions were generated for each of the eight categories and model papers were assembled to illustrate particular attributes for training purposes. Four history raters (three AP history teachers and one history graduate student) were trained in the use of the new system. They spent two days rating the same set of 85 eleventh-grade papers used in the first study. Raters were observed as they scored papers and were queried about their satisfaction with the rating scales and training procedures. Raters had been given the scoring rubric in two forms: an extended, multipaged form with detailed explanations about each score point for use in training; and an outline of the dimensions. It was expected that after the initial training period the raters would use the outline form. However, they chose to continue to refer to the extended form, more rigidly adhering to the rubric than we expected. Raters reported that they could differentiate among categories and that they could also distinguish among criteria for score points (1-5) within each category. Raters were highly satisfied with the scoring categories and claimed to use similar criteria to score papers produced in their own classrooms.

Data from the second round of scoring were then analyzed. Unfortunately, the findings from these ratings did not significantly advance our research goal. Percentage of exact agreement among raters nudged up to about 35%, but alpha coefficients for rater agreement dropped to around .45. Most disappointing were relatively high intercorrelations (in the .80 range) among rating categories. These strong relationships were confirmed by a factor analysis that produced only two factors, one factor consisting solely of the mechanics rating and the other loading all other categories. These disappointing results forced us to regroup intellectually once again. Fortunately, we were able to compare the results from the first set of ratings by the five history teachers with this set of scores, since the identical student papers were read by both groups of history specialists. The categories in our revised rating scale that mostly highly correlated with the overall content quality rating from the first experiment were Historical Context, Breadth of Elaboration, and Depth of Elaboration; these categories were set aside for future exploration.
We so far had investigated the existence of common implicit criteria for content quality ratings, had analyzed the think-aloud protocols of raters, and had noted criteria used in identifying successful student papers. We then had created a comprehensive list of content-relevant elements, had reduced them to a smaller set of categories for feasibility purposes, and had trained a satisfied group of raters. Yet, we had not seemingly made much progress toward our goal. At this point we realized that our entire process had been guided in large measure by what history specialists said they valued and usually focused upon when they graded papers. It became obvious that such descriptions might reasonably be influenced by the raters' desires to appear to be comprehensive and thoughtful—in other words, by the social desirability of their answers.

**Scoring Criteria: Pass Three**

A new strategy for developing scoring criteria was employed, using the model derived from expert-novice comparisons (see Chi & Glaser, 1980, for an illustration). Rather than focus on what experts said they did, we were going to study their actual performance on tasks identical to those provided by the students. Three expert historians who were advanced graduate students in history, three secondary school history teachers, and three Advanced Placement students were asked to write answers to the same essay question used in the study above and to think aloud to permit us to assess their processes. The analyses of the essays produced by this process as well as our analyses of the think-aloud transcripts resulted in some clear direction for us in the area of criteria generation: Our analyses showed that all experts and some teachers used the elements in Table 4 to construct their essays.

<table>
<thead>
<tr>
<th>Table 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elements Used by All Experts and Some Teachers in Essay Construction</td>
</tr>
</tbody>
</table>

- A strong problem or premise that directed a focused answer
- Use of prior knowledge, including principles as well as facts and events for elaboration
- Text references (i.e., Lincoln speech)
- Explicit effort to show interrelationships

In contrast, very bright but relatively inexperienced students and some teachers leaned heavily on the text in two ways. First, they often simply paraphrased or even restated the text in their answer. Second, they tried to cover all elements discussed in the text and were unable to distinguish between more and less important details. As a result of this analysis, a scoring scheme was developed that included all of the elements in Table 4, augmented by an overall general impression score. We were ready for new data collection.

**Rethinking Our Task**

The first major redirection of this project occurred because of the paucity of quality textbooks and resulted in turning this assessment research toward the dual goals of measuring understanding and knowledge acquisition in the context of a particular subject matter corpus. The expert-novice analyses reshaped our focus in a second major way. If we accepted that prior knowledge in subject matter was essential to both premise-driven and elaboration components of quality of understanding, then it was clear that we should design our assessment situations to include explicit supports to enable students to access such information. We believed that we could do this in any number of ways and decided to explore a range of options, details of which we will
expose below. More importantly, we perceived that this decision dramatically revised our view of
assessment. We decided that the assessment situation itself should help students to perform the
best that they could. We had moved into the blurry territory between learning and testing.

Revising the situation. Our next step was to create new questions to relate to the class of
expert behaviors we had proposed as criteria. We decided to have all students read both the
Lincoln and the Douglas texts to permit them to use comparison as a rhetorical structure. We
developed two variations of essay questions, or prompts, which we experimentally crossed: One
treatment condition included a narrative context for the prompt and asked the student to imagine
being in the pre-Civil War period and focus on an imaginary cousin as the audience for the essay;
the other prompt presented the task as a more typical school assignment with the teachers as the
implicit audience. A second set of treatments varied the instructions given to the student to assist
their access to relevant prior knowledge. Although both conditions explicitly directed students to
use their previous understanding and knowledge about the historical period in answering the essay
question, one condition asked a series of stepped, short-answer questions to be completed before
the student began to write the essay (see Table 5).

We also developed a prior knowledge test, basing it on the broad model developed by
Langer (1984), for two purposes. First, we wanted to help students access relevant prior
knowledge; second, we wanted to look at the relationship between that measure and rated use of
prior knowledge in the essay. This 20-item test was created using a set of specifications to control
the nature of the content queried. Students were to write brief descriptions or definitions for each
of the terms provided, some of which were facts and events (e.g., Dred Scott decision), and some
of which were at the principle (or at least concept) level (e.g., sectionalism). A few terms were
irrelevant to the passage, and some were only tangentially relevant.

The new test administration sequence required two days. On the first day the students were
to complete a personal information form (including details about their interests, age, etc.) and the
20 item prior knowledge measure. They then were to read the Lincoln and Douglas text segments
and complete a short (14-item) multiple-choice test on information in the speeches. On the second
day, they were to receive the essay question, write about 45 minutes, and complete a short
debriefing questionnaire that asked for their reactions to the testing and for their estimates of their
performance on the set of tasks tested. Following a pilot test in two Los Angeles classrooms, we
tried the new assessment package in twelve classrooms in Springfield, Illinois. 3

The Illinois Study
The purpose of the Illinois study was to test the assessment procedures under large-scale
assessment conditions and to obtain data to bear upon the validity of our findings. Here we have
space for only a short description and discussion of this study. In brief, 250 students in 11th
grade participated, equally assigned from AP, college preparation, and regular classes. Two full
class periods were allowed for the assessment. Students were told they were participating in a
UCLA study to develop new measures for history. Since there were four treatment variations
(stepped essay prompts/short prompts/narrative context/school context), students received their
packets assigned at random within each classroom. On-site observers from UCLA administered
the materials and collected information from teachers about their views of students' relative
strengths in history, reading, test taking, and writing, and information about each teachers'
instructional efforts in the topic area. In addition, we collected data from transcripts that reported
students' course experience, grade point averages, and standardized test scores in writing, social
studies achievement, and reading comprehension.

3 A good place for prior knowledge on Lincoln and Douglas.


Table 5
Sample Prompt: Step Narrative Version

Topic:

Imagine that it is 1858 and you are an educated citizen living in Illinois. Because you are interested in politics and always keep yourself well informed, you make a special trip to hear Abraham Lincoln and Stephan Douglas debating during their campaigns for the Senate seat representing Illinois.

1) Unlike other tests, we hope you really will try to imagine yourself in the historical period of the debates, so take a couple of minutes to describe yourself, your family, and your work. (Spend about 2-3 minutes.)

2) As a well-informed citizen, you are aware of the many important events, laws, and court decisions that relate to the debates. List as many of these as you can. (Spend about 3-4 minutes.)

3) List, if you can, some principles that underlie our form of government and that are relevant to the debate. (Spend 3-4 minutes.)

4) While listening to the debates, you begin to think about the major problems confronting the nation. Some of these problems relate to principles upon which our government was founded. List the major problems you can think of. (Spend about 3-4 minutes.)

5) After the debates, you return home to find your cousin from England who has come to the U.S. for a visit. Your cousin asks you about some of the problems that are facing the nation at this time. Write the answer that you would give to your cousin, telling him/her about at least two problems that you feel are important. You can write this either like a regular essay or like a story. Just be sure to give your cousin the clearest picture you can. You may use any of the information you've identified above in your answer.

Be sure to describe each of the problems clearly and tell your cousin about events, laws, court decisions, and major principles of U.S. government that are related to the problem. Also explain the different solutions that are proposed to the problem, and give an example of what might happen if these solutions were adopted.

As a conclusion to your paper, write a brief summary that integrates the two problems and states your own position on the whole topic.

To obtain results, prior knowledge scoring rubrics were developed and applied to student responses. Scores ranged from 3, a fully elaborated answer, to 1, an incorrect or incoherent response. Two graduate students were trained to use the prior knowledge rubric and achieved .96 interrater reliability across the total measure (individual item agreements for the 20-item measure ranged between .70 and .96; 15 items had at least .86 agreement and only 2 fell below .80). Essays were rated using the new scoring rubric presented in Table 6.
Table 6
Elements of Cognitively Sensitive Assessment Scoring Rubric

<table>
<thead>
<tr>
<th>Problem Focus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prior Knowledge: Principles and Facts</td>
</tr>
<tr>
<td>Text Reference</td>
</tr>
<tr>
<td>General Impression</td>
</tr>
</tbody>
</table>

This time our empirical results were encouraging. Interrater reliabilities for the essay subscales ranged between .85 and .98. Intercorrelations among subscales were found between .0 and .60, supporting the premise that different aspects of student content quality were being assessed. Our findings also shed some light on the validity of the rubric. First, we determined that the measures reflected the different ability levels of the sample, with AP students scoring twice as high as the slower students on prior knowledge measures and on overall essay scores, and more than three times higher on use of principles in the essay. Our findings also showed strong relationships between teacher judgment of overall student achievement in history and our data (r = .42 for essay, .63 for prior knowledge). Our measures and standardized tests correlated .73 and .43, a variation based upon standardized test content.

Scoring Criteria: Pass Four

We reviewed our findings and decided it was time to test whether regular history teachers could be trained to use the cognitive scoring scheme. We also decided to revise the scale in a number of ways: to add categories for misconceptions and interrelationships, since in our own discussions we had not found a place in our system to take such concerns into account; and to refine the scale points for principle and problem focus. The categories in the scoring rubric re displayed in Table 7.

Table 7
Cognitive Assessment Scoring Rubric (1989)

<table>
<thead>
<tr>
<th>Presence of Problem Focus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prior Knowledge: Principles</td>
</tr>
<tr>
<td>Prior Knowledge: Facts and Events</td>
</tr>
<tr>
<td>Text</td>
</tr>
<tr>
<td>Interrelationships</td>
</tr>
<tr>
<td>Misconceptions</td>
</tr>
<tr>
<td>General Impression</td>
</tr>
</tbody>
</table>

We then conducted a training session with four high school history teachers to test the feasibility of our modified scoring approach. The training took approximately four hours, followed by the scoring session. Once again, we were very encouraged by our results. The prior knowledge measures and the essays were found to be reliably scored by teachers. Slightly lower interrater agreement overall was found for the high school teachers compared to the level obtained by project research assistants (alpha = .93 instead of .96). The interrater reliabilities on the essay
subscales for teachers were in the .80-90 range, except for the newly added misconception category (.68). Correlations between the prior knowledge measure and related elements of the scoring scheme were all reasonably high, averaging around .59, except for misconceptions (-.20) and text material (-.28). We conducted a factor analysis on essay subscales, and two major factors emerged. One factor included overall scores on content quality, the use of principles-based prior knowledge, premise-focused writing, and interrelationships. The second factor included misconceptions, the use of facts, and the use of text-based material. Although we are not completely convinced that this factor structure is sensible, the configuration of elements as it relates to the cognitive construction of meaning (factor one) and of the application of disconnected, and perhaps incorrect, information (factor two) is provocative.

Next Steps

Research subsequent to the Illinois study has been undertaken to verify the utility of the scoring system across topics, age ranges, and test administration conditions. We are looking at the performance of 9th-, 10th-, and 11th-grade students in two school districts. Data have been collected and are presently under analysis. The utility of this approach was tested using two additional assessment topics. Both of these topics are drawn from the pre-Revolutionary War period and include texts by Paine, Henry, and Inglis. In addition, new materials have been developed for an extended assignment that involves Long and Roosevelt texts from the Depression period and incorporates as well additional resource materials for students' optional use. We anticipate a total of five hours will be needed for the assessment.

Limitations and Cautions

We have recounted the details of this effort to provide some insight into how assessment systems might be developed to reflect better the ways students actually learn and integrate subject matter material into their repertoires. We detailed our troubles and dead ends to demonstrate that the process of developing new kinds of useful and valid achievement measures is difficult and time consuming. New approaches to assessment are essential, but their development must be grounded in a theoretical view of learning. Establishing the validity of such new measures is also a difficult proposition. At least three major problems exist. One difficulty is the circular nature of new test development. Measures need to relate to but not be too strongly predicted by existing measurement strategies. A second problem with "deep understanding" tasks is the clear lack of systematic experience for the average student. Most students reported to us that our tasks were unusual for them. Their overall performance levels were exceptionally poor. To determine if our measures are truly valid (that is, if they reflected the desired class of learning), experimental studies must be constructed where students are trained explicitly in the process of integrating specifically presented material with various types of prior knowledge. Third, and most difficult, an optimal level of generality for task descriptions and scoring criteria is needed. This level must be sufficiently detailed to control raters' scoring behavior and to be valid for specific tasks. It must be sufficiently general to provide cues for teachers to use in planning and implementing instruction. A rough approximation of how such information can be economically displayed is provided in the specifications presented in Table 8. Such specifications would be augmented by detailed scoring rubrics with scale point definitions and also by a set of student papers illustrating, on different topics, various levels of proficiency. Clearly, a new program of psychometric research is needed. In the interim, we suggest that validity studies include criterion analyses by experts, experimental training studies, multiple measures of student learning processes, and demonstrations of statistical and conceptual connections to other reasonable estimates of performance, even including standardized tests.
<table>
<thead>
<tr>
<th>Discourse Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Informative writing</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Subgenre</th>
</tr>
</thead>
<tbody>
<tr>
<td>Explain/infer</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Major Cognitive Process</th>
</tr>
</thead>
<tbody>
<tr>
<td>To demonstrate the acquisition of new knowledge or concept by contextualizing and elaborating position using prior knowledge (principles and facts)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Writing Process Measured</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drafting</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Audience</th>
</tr>
</thead>
<tbody>
<tr>
<td>Imaginary, peer</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Topic Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subject matter based:</td>
</tr>
<tr>
<td>History: a summary of major position by opposing statesmen</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Information Given in Prompt</th>
</tr>
</thead>
<tbody>
<tr>
<td>History: Text of speeches or essays written by historical figures, e.g., Lincoln</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Format</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brief text</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Prior knowledge cues: Consisting of appropriate and inappropriate terms for specific processes, facts, or principles</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Amount:</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 or 3 pages (no more than 10 minutes of reading)</td>
</tr>
<tr>
<td>A list of 10 to 20 entries for prior knowledge</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Content:</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Organizing premise</th>
</tr>
</thead>
<tbody>
<tr>
<td>Explicit use of prior knowledge, principles and facts (either provided or student generated) to explain or elaborate</td>
</tr>
</tbody>
</table>
(Table 8 continues)

Avoidance of misconceptions

Structure:

Relevant text references

Show interrelationships using text and prior information

Administrative Conditions

Time: 45-60 minutes

Resources: Students may refer to text and prior knowledge list during essay preparation

Interaction: None

Sample Prompt

Segment of Patrick Henry's speech, plus list of prior knowledge measure

Read the speech taken from the period just before the American Revolution. You are supposed to explain to a cousin visiting from Canada what Patrick Henry meant and what led him to the position he is in. Use help from the list of information to provide a clear answer.

Parallel prompt

Same except pre-Civil War, Stephen Douglas

We know that tests have driven instruction in the past. Can tests of the sort we are developing do so in a productive rather than a destructive way? What evidence do we have that teachers of history focus on the integration of new knowledge with prior information—the view that learners construct meaning? Are such tasks within the capability of all students? When we are constantly bombarded with stories that students don't know where the Pacific Ocean is or the half-century in which World War I occurred, is it naive to think that they can accumulate knowledge and use it to make inferences and explanations. These questions must be pursued. We believe that there are specific next steps to be accomplished. A major challenge is the development of a new theory of test design and validation, one that emphasizes individual learning rather than individual differences. Test designers must recognize that the measurement of significant processes takes significant time, and consequently tests of many short items and broad content sampling may need to be supplanted or supplemented by fewer more complex assessment situations. We need to develop concepts that will allow teachers to understand how to use such measures as an integral part of their instruction. Finally, we must get ready for the serious task of educating policy makers and the public about new models of assessment. We must counsel patience and anticipate that results are going to look worse, especially with new challenging measurement approaches, before they look better. When improvements eventually occur on cognitive measures such as those we have explored, we want them to reflect real and trustworthy learning for all students.
References


PART III

GUIDELINES FOR USING CRESST'S
COGNITIVELY SENSITIVE ASSESSMENT SYSTEM
PART III: GUIDELINES FOR USING CRESST'S COGNITIVELY SENSITIVE ASSESSMENT SYSTEM

Our approach to assessing students' deep understanding of content knowledge involves the following primary components:

• assessment (and activation) of students' general and topic-relevant prior knowledge (referred to below as the prior knowledge test),

• provision of primary source materials, i.e., new information in written text for students to read (text material), and

• assessment of students' content understanding revealed by their writing in response to a contextualized prompt calling for integration of prior and new knowledge to explain subject matter issues (writing task and essay scoring rubric).

The rest of this section discusses the characteristics of these elements, particularly the rubric for scoring student writing, and the procedures for using these materials to collect data.

Assessment Task Characteristics

Prior knowledge test. This measure consists of 20 short answer items on concepts and facts related to the text material, to the general topic (historical period) in question, and to basic principles and concepts that recur in the content area (American historical thought). Each item is scored on a 0-4 point scale, with "1" indicating the student has missed the point and "4" indicating high understanding of the term. Our research suggests that the administration of a short prior knowledge test is very desirable if students are to demonstrate deep understanding and the ability to connect new information (from the just-read texts) to ideas, concepts, and facts they have already learned. Giving a prior knowledge test before having students read new text material activates their existing relevant knowledge, helps them comprehend new information, and helps them synthesize its meaning in the light of previous learning. In addition, it provides a measure of the background knowledge students may be able to bring to bear on the writing task. These tests have scoring keys and tend to be very reliable (see Part II).

Text materials. The provision of primary source materials for students to read and incorporate in their writing is an important element of our assessment task. These texts have been excerpted from primary source materials, typically speeches or letters for the history topics and descriptions of experiments for the science topics. Students have been given time to read two opposing views on the same topic, such as a Lincoln-Douglas debate, and their essays use these readings as a point of departure. In the development and validation of our materials, we included short multiple choice reading comprehension measures after students read the text material, but we do not believe these are essential to the assessment process. We recommend their inclusion only if literal comprehension is regarded as a principal educational goal.

Writing task. The most basic element of our assessment approach is the student writing task. In our sample materials presented in this handbook the writing task is an essay written during one class period by the student without help from the teacher or peers. However, the task could be a simple written homework assignment or a longer paper and might even be a cooperatively developed product rather than the product of an individual student. Our research suggests that students' thinking and writing benefit from long composing periods where possible. Our scoring scheme will work (with some adjustments—noted later) for a variety of assessment conditions.

The essay prompt asks students to explain some important issues using information in the text and knowledge that they have already learned in school or elsewhere. Our research suggests that the essay prompt should be highly contextualized to help students reveal their understanding of the issues they are to address. We tried various kinds of prompts, some with a strong narrative context e.g., "You are a farmer in Illinois in the 1850's..." as well as those which ask students to
write a specific explanation to another person, e.g., "You want to explain to your cousin..." Although we found no significant improvement for using such prompts, they seemed to enhance student fluency. We recommend contextual prompts to heighten student understanding of the task and to provide an audience (even though we understand that school-based writing has only limited "real life" relevance) as this is becoming common practice in literature-based writing programs.

The CREST essay scoring rubric. We employ a criterion-referenced scoring strategy: papers are judged in terms of prespecified standards rather than in relationship to one another. This method implies that the "best" papers in a given group will not receive the highest rating if they fall short of explicated standards of performance.

The Scoring Rubric for the essay consists of five scales or dimensions:
1) General Impression of Content Quality;
2) Number of Principles or Concepts;
3) Incorporation of Prior Knowledge of Facts and Events; and
4) Proportion of Essay Using Text-Based Detail;
5) Misconceptions.

General Impression of Content Quality means the overall competence the student exhibits in answering the question. This dimension focuses on the way students demonstrate their understanding of the concepts (and in history, the historical era discussed). Number of Principles is a count of the major principles or concepts that students use with comprehension in their essay. Incorporation of Prior Knowledge is a judgment about the amount of relevant information that is not explicitly in the primary text material, which the student uses to illustrate or support his or her perspective. Proportion of Essay Using Text-Based Detail is an assessment of how much information presented in the essay is derived from primary text materials. Misconceptions asks raters to judge the extent to which students reveal major misunderstandings of concepts or of combinations of facts. A complete guide to the Essay Scoring Rubric is in Part VII of this handbook.

Each essay is scored on all five dimensions using a scale ranging from 0 to 5. Our studies indicate that the best essays receive low scores on Misconceptions (i.e., contain few misconceptions) and low-to-moderate scores for Proportion of Essay Using Text-Based Detail, with high scores on the other scales.

Over a period of years we examined a number of possible scales and dropped some because they provided little information or were inefficient. For example, the scoring rubric originally included an overall or holistic rating, General Impression of Expression, related to the student's ability to use standards of written English and to communicate well. As this scale correlated highly with General Impression of Content Quality, it technically added no information and was dropped from the rubric. However, we have found that our raters, who are typically drawn from the ranks of secondary school teachers, feel it is important to include some rating of language. In some settings, this General Impression of Expression scale could be included for instructional feedback purposes. Similarly, the rubric once contained a scale which rated the amount of the essay organized around concepts, Concept Organization. This scale was dropped because it correlated highly with Number of Principles or Concepts, and the latter was easier for raters to score reliably. However, for classroom level use, some teachers may be more satisfied rating an essay for its use of organizing principles than merely counting the number of concepts or principles present in the essay. Proportion of Essay Using Text-Based Detail is a sensible dimension only if the assessment task follows our model of providing reading materials as the context for the essay writing. Otherwise, this scale should be deleted.

The scales of the scoring rubric are differentially useful. They may be weighted differently depending upon the particular objectives of the assessment. For example, if reading comprehension or learning from new materials is important, then there may be a reason to look for high scores on the Proportion of Essay Using Text-Based Detail dimension. When modifying the scoring rubric, directions to students should be revised to cue them as to what is important for a particular assessment.

High levels of interrater reliability have been achieved for the scoring rubric across different topics and different raters (see Part II). The lowest reliability estimates have been associated with
the dimension of *Misconceptions* (correlation of .68). Relatively high levels of agreement are essential for the use of the scale with validity. They also permit the use of single raters when the purpose of the assessment is to determine the overall standing of a group or to provide diagnostic feedback on instruction. We still recommend the double scoring of some portion of the essays, e.g., 20% of the papers would be a recommended minimum, as a check on rater agreement. High levels of agreement can only be obtained when careful rater training has been conducted.

**Procedures for Collecting Data**

**Schedule of tasks.** In a typical (non-experimental) assessment situation, the student tasks and minimum times required to administer the sample materials provided in Part IV of this handbook are as follows:

- 15 minutes to complete prior knowledge test
- 25 minutes to read primary source text
- 50 minutes to read prompt and write essay.

Of course, students may be given more time to read the text materials, particularly if their reading level is quite low. In our experience, students need at least 45 minutes to write essays in response to the prompts provided here and could benefit from an even longer period if possible.

Longer reading and writing tasks may be used instead of those provided here. For example, in one study we allowed students three days (not just three class periods) in which to read as many text materials as they liked from a selection of twenty, to take relevant notes, and to write an essay in response to a prompt.

The basic tasks and schedule for teachers include the following:

- 10 minutes to read introductory letter and description of tasks
- 10 minutes to read and learn the instructions (script) for testing and familiarize oneself with the measures
- 60 minutes to administer the student measures (prior knowledge test, reading, and essay)
- 10 minutes to complete the teacher questionnaire (optional)
- 10 minutes to complete the prior knowledge test (optional).

You will need to revise the introductory letter and instructions to the teacher provided in the Teacher Packet in Part IV according to your local situation. The teacher questionnaire, covering the teacher's background and instructional practices, was an empirical tool for us but may provide you with ideas for alternative indicators that you may wish to measure and relate to student performance on the prior knowledge test and essay task.

We asked teachers to complete the prior knowledge test in order to obtain multiple examples of answers that would indicate the highest level of understanding, a "4" on our scale, for our use in rater training. You may wish to do the same, particularly if you create new topics to assess. If so, you may need to take steps to reassure teachers that they will not be judged in any way on the basis of their answers. Asking teachers to volunteer to "take the test" may solve the problem.

**Organizing student packets.** If you use more than one prompt, you will need to randomly assign prompts to students in each classroom and grade level assessed. Packets of student materials should be organized ahead of time, with students' names, ID numbers, teacher, school, and prompt version typed on a label affixed to the front of their packets to facilitate management of packets during the assessment period. Color-coded materials simplify administration procedures as well. Note that critical identifying information (such as student name and/or ID, school, teacher) should be written by students at the top of every prior knowledge test and essay so that these pieces can be identified later when they have been removed from the student packets for scoring.
Organizing teacher packets. Each teacher should receive a packet of materials that includes: an introductory letter explaining the purpose and general procedures of the assessment, a schedule and description of the tasks, directions or script for administration of the measures, and additional materials as appropriate (e.g. teacher background and instructional practices questionnaire, and prior knowledge test to complete). See Part IV for a sample teacher packet. Teachers should be given these materials well prior to the time of assessment so that they may become familiar with the procedures and purposes and have time to ask questions.
PART IV

SAMPLE DATA COLLECTION MATERIALS
FOR STUDENTS AND TEACHERS
PART IV: SAMPLE ASSESSMENT MATERIALS
FOR TEACHERS AND STUDENTS
(U.S. HISTORY AND CHEMISTRY TOPICS)

This section of the handbook contains a complete set of teacher materials and a set of student materials for assessing student understanding of one U.S. history topic (the Civil War era). It also contains student materials for assessing one chemistry topic (analysis of an unknown substance). You may duplicate and use these materials as is or you may use them as models for developing your own assessment program.
1. Teacher's Packet

   Introduction Letter
   Description of Tasks
   Instructions for Testing
   Teacher Questionnaire
November 3, 1989

Dear Colleague:

Thank you very much for assisting us with our project to develop new ways to measure students' deep understanding of social studies. The essays and test results provided by your classes will help us refine our measures to effectively elicit what students know.

The two days of testing are scheduled for Thursday and Friday, Nov. 16 and 17. We have enclosed a description of the tasks that students will complete on those days. We hope you will not do any specific teaching for our study and not discuss the tasks with the students ahead of time. They should only be told that they will be participating in a project that focuses on their understanding of history.

In our experience, students are more involved and more likely to demonstrate what they actually know if they feel that the tasks are a regular part of school, are important, and will be reviewed by you (rather than just part of a research study for someone they will never see again). We would like you to emphasize that it is important for them to put in sincere effort. Since many students will have limited knowledge of the pre-Civil War period this early in the year, you might reassure them that you will take into consideration what material has and has not been covered in class prior to the testing. We will send you copies of the essays your students write as well as their scores on the other measures, for your own use.

On the day before the testing (Nov. 15), we will deliver our materials to your school, including a script for you with specific instructions for the two days. During the testing we will have one staff person in each classroom to serve primarily as an observer. If the students raise any questions that you cannot answer on the basis of the information we have provided, the staff person will be able to respond.

During the two days of testing there will be some time when students are reading or writing during which we would like you to complete some brief forms about your students' abilities in history and the type of instruction and testing to which they are accustomed.

We plan to report data so that your identity is protected and our findings will be used only for the development of these measures.

Please have available a seat assignment to occupy those students who finish early each day.

Thank you again for your participation in our project. We look forward to working with you.

Cordially,

Eva L. Baker
Director
Description of Tasks

The students will receive a packet that contains everything they will need for the two days except a pen. Please do not discuss these tasks with the students ahead of time. Three different versions of materials will be divided among the students in the class, so you should expect to see some differences in the test packets.

DAY 1

1) Prior knowledge test in U.S. History: This test consists of a list of terms or words for which the students will write definitions or given examples.
2) The primary source materials: These are the provided texts to be read by the students (e.g., edited versions of a speech by Douglas and speech by Lincoln).
3) Reading comprehension test: This is a multiple choice test of literal and low inference comprehension of the reading texts.*

DAY 2

1) Writing prompt: Students will read directions and essay prompt.
2) Writing: Students will write their essays on paper provided.
3) Attitude measuring: An questionnaire asks students about the way they felt while writing.*
4) Debriefing questionnaire: A few short questions asks students about their reactions to the testing experience.

*The procedures are optional.
Instructions for Testing

The instructions in bold-face are to be read aloud. The plain faced type is directions for the teacher administering the test.

DAY 1 - Nov. 16

"We are going to try to assess your understanding of history using a variety of approaches. Instead of asking you to do only one thing to show your understanding, today you will be asked to do two tasks: to try to remember some ideas you already learned, and then to read some new material and show how much you learned from it. Tomorrow you will be asked to combine the understanding you already have shown by imagining you are in the same historical period as your readings and to help solve an information problem. Some of this may be easy for you and some unfamiliar. But it is important that you try your hardest to show what you know. I will be looking at your work and taking it into account."

Pass out packets

"Check to see that the name at the top of your envelope is yours. If your name is not on the label, carefully print it there now. If my (teacher's) name is not already on the label, carefully print it also."

"Open the packet and take out the blue set of papers. Do not remove any other papers."

"You will have about 15 minutes to answer the questions on these pages. Be sure to write clearly. This is a test of what you know about some aspects of U.S. History. If you finish early, check your answers and wait until I tell you to go on. You may begin now."

Wait 15 minutes unless you observe that everyone has finished earlier in which case you should go on.

"Time is up. If you have not already finished, please draw a line across the paper to show where you stopped. Now put the blue pages back in the envelope and take out the white papers. You have about 25 minutes to read these materials. You may underline or make notes on the text if you want to." If you finish early, you may read it again for better understanding.

Wait about 25 minutes. Check to make sure that students who finish are not going on.

"Time is up. If you have not already finished, please draw a line across the text to show where you stopped. Now put away the white pages and take out the pink set titled Information Measure and answer these questions on what you just read. You may not look back at the reading materials. You have the rest of the period." (about 10 minutes)
"When you are finished, be sure your name and your 3-digit number (see label on envelope) is on the two tests you took. Put everything back in your envelope at the end of the period. Be sure your name is on the label on the outside of the envelope. Please pass your envelope to the front of the room. If you finish early, work on your seat assignment."

End first day
DAY 2 - Nov. 17

"Today we are going to be finishing up the project that we started yesterday."

pass out packets

"Open your envelopes and take out the white reading passages from yesterday, the yellow paper, and the lined writing paper. Today you will be doing some writing based partly on what you read yesterday. Feel free to look back at the reading passages. There are several different versions of the materials in the room - so don't worry if your papers look different from your neighbor's.

Let the students read the prompt (about 2 minutes). Then say:

It is very important that you explain the most important ideas and issues to your listener. Your essay should be based on two major sources:

1) the general concepts and specific facts you know about American History and especially what you know about the history of the specific period;
2) what you have learned from the readings yesterday

Be sure to show the relationships among your ideas and facts.

You should use the blank paper for writing your essay. You'll have about 45 minutes. Be sure to put your name and ID number on every page. Use pen, margins and skip lines. You can start now.

At 15 minutes before the end of the period -

"You have five more minutes - you should be finishing up."
(or you can announce at intervals as you usually do, but be sure to leave 10 minutes at the end for the other measures)

At 10 minutes before the end of the period -

"Stop writing now and check that your name and number are on every sheet of paper you used, even sheets that only have notes on them."

Pause to make sure that everyone is doing this

"Now take out the green questionnaire. Put your name and number on it, and complete it."

At the end of the period -

"Put everything back in your envelope including your essay and any notes you made. [You can keep the UCLA pencil that came in your envelope.]"
Teacher Questionnaire

Name _______________________
School _______________________

Complete one of these forms for each class of yours that participated in our study.

Period during which this class meets: _______________________

Track of this class: regular modified honors AP

Circle the letter of your answer for each question below.

1. Overall, how well do you think your students performed on the tests given over these two days?
   a. very well
   b. pretty well
   c. okay
   d. not very well
   e. not well at all

2. Overall, how well do you think your students performed on the tests compared to the students you taught 3 years ago?
   a. current students performed much better
   b. current students performed somewhat better
   c. current students performed about the same
   d. current students performed somewhat worse
   e. current students performed much worse

3. How well do you think your current students performed on the first test with 20 short answer items?
   a. very well
   b. pretty well
   c. okay
   d. not very well
   e. not well at all

4. How well do you think your current students performed on the multiple choice test about the reading passages?
   a. very well
   b. pretty well
   c. okay
   d. not very well
   e. not well at all
5. How well do you think they performed on the essay?
   a. very well
   b. pretty well
   c. okay
   d. not very well
   e. not well at all

6. If you feel that many of your students didn't perform well on these tests, to what do you ascribe this? (check all that apply)
   a. reading passages were too difficult
   b. students had little prior knowledge of these topics
   c. students had little experience writing essays
   d. other (describe)_____

7. How well do you think these tests elicited what your students know about these subjects?
   a. very well
   b. pretty well
   c. not very well
   d. not well at all

8. How hard do you think your students tried on these tests?
   a. very hard
   b. pretty hard
   c. not hard
   d. didn't really try at all

9. Compared to how hard they usually try on the tests you give, how hard did they try on these tests?
   a. much harder on these tests
   b. a little harder on these tests
   c. about the same on these tests
   d. a little less hard on these tests
   e. much less hard on these tests

10. Compared to homework, how hard did they try on these tests?
    a. much harder on these tests
    b. a little harder on these tests
    c. about the same on these tests
    d. a little less hard on these tests
    e. much less hard on these tests

11. Compared to regular class discussion, how hard did they try on these tests?
    a. much harder on these tests
    b. a little harder on these tests
    c. about the same on these tests
    d. a little less hard on these tests
    e. much less hard on these tests
12. Compared to standardized tests, how hard did they try on these tests?
   a. much harder on these tests
   b. a little harder on these tests
   c. about the same on these tests
   d. a little less hard on these tests
   e. much less hard on these tests

13. How many essays will your students write for this class this year?  

14. How often do your students do the following?

<table>
<thead>
<tr>
<th>Activity</th>
<th>Almost every day</th>
<th>Once or twice a week</th>
<th>Once or twice a month</th>
<th>Once or twice a semester</th>
<th>Never</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Read primary materials (such as Federalist Papers, Brown vs. Board of Education)</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>b. Read novels, plays, essays, or poems related to class</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>c. Read material from textbook in class</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>d. Engage in simulations, role-playing, or acting things out</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>e. Write research papers</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>f. Give oral reports</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>

15. How much homework do you typically assign each day in this class?
   a. none
   b. half an hour
   c. 45 minutes
   d. 1 hour
   e. more than 1 hour

16. How often do you administer quizzes or tests in this class?
   a. once a week or more
   b. several times a semester
   c. once a semester
17. Taking all the tests and quizzes together that you give in this class all year, what proportion of TOTAL CREDIT do you allocate to the following types of questions?

a. ___% multiple choice, true/false, sentence completion
b. ___% short answer (less than 1 page)
c. ___% essay
d. ___% other

18. Do you think that a week-long group task in which your students would learn new information and would write a paper integrating the new knowledge with their prior knowledge of history would be a worthwhile experience for them?

a. very worthwhile
b. somewhat worthwhile
c. not very worthwhile
d. a waste of time

19. Would such a task provide useful feedback to you as a teacher?

a. very useful
b. somewhat useful
c. not very useful
d. a waste of time
The topics listed below are among those that might be included in a U.S. History course. For each topic, please answer the questions for Sections A and B below.

For Section A: What percentage of the students in this class would you expect to have some knowledge of this topic?

For Section B: In this course, how do you treat this topic?

<table>
<thead>
<tr>
<th>TOPICS</th>
<th>A. % of class with some knowledge of topic</th>
<th>B.</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1. deflation</td>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. abolitionists</td>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. states' rights</td>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. New Deal</td>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Stamp Act</td>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. recession</td>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. imperialism</td>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Federalism</td>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. gold rush</td>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. margin buying</td>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. communism</td>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. Black Thursday</td>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13. Articles of Confederation</td>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14. conservation</td>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15. Missouri Compromise</td>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16. dust bowl</td>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17. industrialization</td>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18. balance of power</td>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>19. court-packing plan</td>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20. Reconstruction Finance Corporation (RFC)</td>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Write below any of your objectives for this class that you feel match or relate to the tasks your students completed during the past two days:

Please record any suggestions you have for improving our methods and materials:
2. Student's Packet: American History/Civil War

Prior Knowledge Test
Texts
Essay Prompt
Student Questionnaire
How Much Do You Know About U.S. History?

Directions: This is a list of terms related to U.S. History. Most of them are related to the period of the Civil War, but some of them are from other periods in U.S. History.

In the space after each term, write down what comes to mind when you think of that term in the context of U.S. History. A brief definition would be acceptable, or a brief explanation of why that person, place, or thing was important. If the term is general, such as "Civil rights," give both a general definition and a specific example of how the term fits into U.S. History, if you can.

Good Example: CIVIL RIGHTS. Rights guaranteed to all citizens regardless of race, sex, religion, etc. Blacks fought for their civil rights in the 1960s. Martin Luther King, Montgomery bus boycott.

Do not define the term by simply restating the same words.

Bad Example: SURVIVAL OF THE FITTEST -- only the fittest survive.

Even if you are not sure about your answer, but think you know something, feel free to guess. There are probably more items here than you will be able to answer in the time given. Start with the ones you know best, and work quickly so that you can answer as many as possible. Then go back and answer the ones of which you are less sure. Do not spend too much time on one specific item.

1. popular sovereignty

2. Dred Scott

3. Communism

4. Missouri Compromise

5. industrialization

6. Gold Rush

7. bleeding Kansas
8. states' rights

9. Federalism

10. underground railroad

11. imperialism

12. Whig

13. Kansas-Nebraska Act

14. Abolitionists

15. sectionalism

16. westward movement

17. constitutionality

18. New Deal

19. party platform

20. balance of power
Texts: Lincoln-Douglas Debate

Directions: As Abraham Lincoln and Stephan Douglas campaigned for the office of Senator from the state of Illinois, they held seven joint debates throughout the state. Read the following passages to understand as well as possible what Lincoln and Douglas discussed in one of their debates.

Stephen A. Douglas*

Mr. Lincoln tells you, in his speech made at Springfield, before the Convention which gave him his unanimous nomination, that—

"A house divided against itself cannot stand."

"I believe this government cannot endure permanently, half slave and half free."

"I do not expect the Union to be dissolved, I don’t expect the house to fall; but I do expect it will cease to be divided."

"It will become all one thing or all the other."

That is the fundamental principle upon which he sets out in this campaign. Well, I do not suppose you will believe one word of it when you come to examine it carefully, and see its consequences. Although the Republic has existed from 1789 to this day, divided into Free States and Slave States, yet we are told that in the future it cannot endure unless they shall become all free or all slave. For that reason, he says, that they must be all free. He wishes to go to the Senate of the United States in order to carry out that line of public policy, which will compel all the States in the South to become free. How is he going to do it? Has Congress any power over the subject of slavery in Kentucky, or Virginia, or any other State of this Union? You convince the South that they must either establish slavery in Illinois, and in every other Free State, or submit to its abolition in every Southern State, and you invite them to make a warfare upon the Northern States in order to establish slavery, for the sake of perpetuating it at home. Thus, Mr. Lincoln invites, by his proposition, a war of sections, a war between Illinois and Kentucky, a war between the Free States and the Slave States, a war between the North and the South, for the purpose of either exterminating slavery in every Southern State, or planting it in every Northern State. He tells you that the safety of this Republic, that the existence of this Union, depends upon that warfare being carried on until one section or the other shall be entirely subdued. The States must all be free or slave, for a house divided against itself cannot stand. That is Mr. Lincoln’s argument upon that question. My friends, is it possible to preserve peace between the North and the South if such a doctrine shall prevail in either section of the Union? Each of these States is sovereign under the Constitution; and if we wish to preserve our liberties, the reserved rights and sovereignty of each and every State must be maintained. I have said on a former occasion, and I here repeat, that it is neither desirable nor possible to establish uniformity in the local and domestic institutions of all the States of this Confederacy. And why? Because the Constitution of the United States rests upon the right of every State to decide all its local and domestic institutions for itself. It is not possible, therefore, to make them conform to each other, unless we subvert the Constitution of the United States. Our safety, our liberty, depends upon preserving the Constitution of the United States as our fathers made it, inviolate, at the same time maintaining the reserved rights and the sovereignty of each State over its local and domestic institutions, against Federal authority, or any outside interference.

The difference between Mr. Lincoln and myself upon this point is, that he goes for a combination of the Northern States, or the organization of a sectional political party in the Free States, to make war until they shall all be subdued, and made to conform to such rules as the North shall dictate to them. His answer to this point, which I have been arguing, is, that he never did

* From Political Debates Between Abraham Lincoln and Stephen A. Douglas (Cleveland, 1902), pp. 43-47.
mean, and that I ought to know that he never intended to convey the idea, that he wished the people of the Free States to enter into the Southern States and interfere with slavery. Well, I never did suppose that he ever dreamed of entering into Kentucky to make war upon her institutions; nor will any Abolitionist ever enter into Kentucky to wage such war. Their mode of making war is not to enter into those States where slavery exists, and there interfere, and render themselves responsible for the consequences. Oh, no! They stand on this side of the Ohio River and shoot across. They stand in Bloomington, and shake their fists at the people of Lexington; they threaten South Carolina from Chicago. And they call that bravery! But they are very particular, as Mr. Lincoln says, not to enter into those States for the purpose of interfering with the institution of slavery there. I am not only opposed to entering into the Slave States, for the purpose of interfering with their institutions, but I am opposed to a sectional agitation to control the institutions of other States. I am opposed to organizing a sectional party, which appeals to Northern pride, and Northern passion and prejudice, against Southern institutions, thus stirring up ill-feeling and hot blood between brethren of the same Republic.

I ask Mr. Lincoln how it is that he proposes ultimately to bring about this uniformity in each and all the States of the Union. Does he intend to introduce a bill to abolish slavery in Kentucky? How is he to accomplish what he professes must be done in order to save the Union? There is but one possible mode which I can see, and perhaps Mr. Lincoln intends to pursue it; that is, to introduce a proposition into the Senate to change the Constitution of the United States, in order that all the State Legislatures may be abolished, State sovereignty blotted out, and the power conferred upon Congress to make local laws and establish the domestic institutions and police regulations uniformly throughout the United States. Whenever you shall have blotted out the State Legislatures, and consolidated all the power in the Federal Government, you will have established a consolidated empire as destructive to the liberties of the people and the rights of the citizen as that of Austria, or Russia, or any other despotism that rests upon the necks of the people.

There is but one possible way in which slavery can be abolished, and that is by leaving a State, according to the principle of the Kansas-Nebraska bill, perfectly free to form and regulate its institutions in its own way. That was the principle upon which this Republic was founded, and it is under the operation of that principle that we have been able to preserve the Union thus far. Under its operations, slavery disappeared from New Hampshire, from Rhode Island, from Connecticut, from New York, from New Jersey, from Pennsylvania, from six of the twelve original slaveholding States; and this gradual system of emancipation went on quietly, peacefully, and steadily, so long as we in the free States minded our own business and left our neighbors alone. But the moment the abolition societies were organized throughout the North, preaching a violent crusade against slavery in the Southern States, this combination necessarily caused a counter-combination in the South, and a sectional line was drawn which was a barrier to any further emancipation. Bear in mind that emancipation has not taken place in any one State since the Free-soil party was organized as a political party in this country. And yet Mr. Lincoln, in view of these historical facts, proposes to keep up his electoral agitation, band all the Northern States together in one political party, elect a President by Northern votes alone, and then, of course, make a cabinet composed of Northern men, and administer the government by Northern men only, denying all the Southern States of this Union any participation in the administration of their affairs whatsoever.
Abraham Lincoln

Judge Douglas made two points upon my recent speech at Springfield. He says they are to be the issues of this campaign. The first one of these points he bases upon the language in a speech which I delivered at Springfield which I believe I can quote correctly from memory. I said there that "we are now far into the fifth year since a policy was instituted for the avowed object, and with the confident promise, of putting an end to slavery agitation; under the operation of that policy, that agitation had not only not ceased, but had constantly augmented." "I believe it will not cease until a crisis shall have been reached and passed. 'A house divided against itself cannot stand.' I believe this Government cannot endure permanently, half slave and half free." "I do not expect the Union to be dissolved" -- I am quoting from my speech -- "I do not expect the house to fall, but I do expect it will cease to be divided. It will become all one thing or the other. Either the opponents of slavery will arrest the spread of it and place it where the public mind shall rest, in the belief that it is in the course of ultimate extinction, or its advocates will push it forward until it shall become alike lawful in all the States, North as well as South."

What is the paragraph? In this paragraph, Judge Douglas thinks he discovers great political heresy. I want your attention particularly to what he has inferred from it. He says I am in favor of making all the States of this Union uniform in all their internal regulations; that in all their domestic concerns I am in favor of making them entirely uniform. He says that I am in favor of making war by the North upon the South for the extinction of slavery; that I am also in favor of inviting (as he expresses it) the South to a war upon the North for the purpose of nationalizing slavery. Now, it is singular enough, if you will carefully read that passage over, that I did not say that I was in favor of anything in it. I only said what I expected would take place. I made prediction only -- it may have been a foolish one, perhaps. I did not even say that I desired that slavery should be put in course of ultimate extinction. I do say so now, however, so there need be no longer any difficulty about that. It may be written down in the great speech.

I am not, in the first place, unaware that this Government has endured eighty-two years half slave and half free. I know that. I believe it has endured because during all that time, until the introduction of the Nebraska bill, the public mind did rest all the time in the belief that slavery was in course of ultimate extinction. I have always hated slavery, I think, as much as any Abolitionist - - I have been an Old Line Whig -- I have always hated it; but I have always been quiet about it until this new era of the introduction of the Nebraska bill began. I always believed that everybody was against it, and that it was in course of ultimate extinction.

The adoption of the Constitution and its attendant history led the people to believe so; and that such was the belief of the framers of the Constitution itself, why did those old men, about the time of the adoption of the Constitution, decree that slavery should not go into the new Territory, where it had not already gone? Why declare that within twenty years the African Slave Trade, by which slaves are supplied, might be cut off by Congress? Why were all these acts? What were they but a clear indication that the framers of the Constitution intended and expected the ultimate extinction of that institution? And now, when I say, as I said in my speech, that Judge Douglas has quoted from, when I say that I think the opponents of slavery will resist the farther spread of it, and place it where the public mind shall rest with the belief that is in course of ultimate extinction, I only mean to say that they will place it where the founders of this Government originally placed it.

I have said a hundred times, and I have now no inclination to take it back, that I believe there is no right, and ought to be no inclination, in the people of the Free States to enter into the Slave States, and in with the question of slavery at all.

So much, then, for the inference that Judge Douglas draws, that I am in favor of setting the sections at war with one another. I know that I never meant any such thing, and I believe that no fair mind can infer any such thing from anything I have ever said.

Now, in relation to his inference that I am in favor of a general consolidation of all the local institutions of the various States. I have said, very many times, in Judge Douglas's hearing, that no man believed more than I in the principle of self-government; that it lies at the bottom of all my ideas of just government, from beginning to end. I think that I have said it in your hearing, that I
believe each individual is naturally entitled to do as he pleases with himself and the fruit of this labor, so far as it in no wise interferes with the right of no other State, and that the General Government, upon principle, has no right to interfere with anything other than that general class of things that does concern the whole. I have said that at all time. I have said, as illustrations, that I do not believe in the right of Illinois to interfere with the cranberry laws of Indiana, the oyster laws of Virginia, or the liquor laws of Maine.

How is it, then, that Judge Douglas infers, because I hope to see slavery put where the public mind shall rest in belief that it is in the course of ultimate extinction, that I am in favor of Illinois going over and interfering with the cranberry laws of Indiana? What can authorize him to draw any such inference? I suppose there might be one thing that at least enabled him to draw such an inference that would not be true with me or many others, that is, because he looks upon all this matter of slavery as an exceedingly little thing, -- this matter of keeping one-sixth of the population of the whole nation in a state of oppression and tyranny unequaled in the world. He looks upon it as being an exceedingly little thing, -- only equal to the question of the cranberry laws of Indiana; as something having no moral question in it; so little and so small a thing that he concludes, if I could desire that if anything should be done to bring about the ultimate extinction of that little thing, I must be in favor of bringing about an amalgamation of all the other little things in the Union. Now, it so happens -- and there, I presume, is the foundation of this mistake -- that the Judge thinks thus; and it so happens that there is a vast portion of the American people that do not look upon that matter as being this very little thing. They look upon it as a vast moral evil; they can prove it as such by the writing of those who gave us the blessings of liberty which we enjoy, and that they so looked upon it, and not as an evil merely confining itself to the States where it is situated; and... we agree that, by the Constitution we assented to, in the States where it is exists, we have no right to interfere with it, because it is in the Constitution; and we are by both duty and inclination to stick by that Constitution, in all its letter and spirit, from beginning to end.
Writing Assignment

Imagine that it is 1858 and you are an educated citizen living in Illinois. Because you are interested in politics and always keep yourself well-informed, you make a special trip to hear Abraham Lincoln and Stephan Douglas debating during their campaigns for the Senate seat representing Illinois. After the debates you return home, where your cousin asks you about some of the problems that are facing the nation at this time.

Write an essay in which you explain the most important ideas and issues your cousin should understand. Your essay should be based on two major sources: (1) the general concepts and specific facts you know about American History, and especially what you know about the history of the Civil War; (2) what you have learned from the readings yesterday.

Be sure to show the relationships among your ideas and facts.
Student Questionnaire

This two day testing period tried out some new ways to see what students know about history. Please give us your honest feelings on the questions below. We will keep your answers confidential and your teacher will not see them.

Circle the letter of the best answer for each question below.

1. Which word best describes the time you have spent on this project? Choose one.
   a. interesting
   b. tiring
   c. confusing
   d. boring

2. Which part of the testing did you like best?
   a. the first test: 20 short answer items
   b. the second test: multiple choice test on the reading passages
   c. semantic mapping
   d. the last test: the essay
   e. the whole thing

3. Which part did you like least?
   a. the first test: 20 short answer items
   b. the second test: multiple choice test on the reading passages
   c. semantic mapping
   d. the last test: the essay
   e. the whole thing

4. Overall, how was your performance on the tasks over these two days?
   a. very good
   b. pretty good
   c. okay
   d. not very good
   e. pretty bad

5. On the first test with 20 short answer items, how was your performance?
   a. very good
   b. pretty good
   c. okay
   d. not very good
   e. pretty bad

6. On the multiple choice test about the reading passages, how was your performance?
   a. very good
   b. pretty good
   c. okay
   d. not very good
   e. pretty bad
7. How was your performance on the essay?
   a. very good
   b. pretty good
   c. okay
   d. not very good
   e. pretty bad

8. Which of these statements best describes how you wrote the essay?
   a. I was able to put together new ideas from the reading with other facts and ideas I already knew.
   b. I focused mostly on writing style
   c. I used yesterday's reading passages for most of my ideas
   d. I mostly used information I already knew before I read the passages.

9. Where did you learn most of the information you used on the 20-item short answer test?
   a. the teacher's lectures
   b. the textbook for this class
   c. other classes I've had
   d. televisions
   e. other reading outside of school
   f. I don't know

10. How well do you think these tests show what you know about this subject?
    a. very well
    b. pretty well
    c. not very well
    d. not well at all

11. How hard did you try on these tasks?
    a. very well
    b. pretty well
    c. not very well
    d. not well at all

12. Compared to how you usually try on tests your teacher gives, how hard did you try on these tasks?
    a. much harder on these tests
    b. a little harder on these tests
    c. about the same on these tests
    d. a little less hard on these tests
    e. much less hard on these tests

13. Compared to standardized tests you take once a year or so, how hard did you try on these tasks?
    a. much harder on these tests
    b. a little harder on these tests
    c. about the same on these tests
    d. a little less hard on these tests
    e. much less hard on these tests
14. Compared to homework,
   a. much harder on these tests
   b. a little harder on these tests
   c. about the same on these tests
   d. a little less hard on these tests
   e. much less hard on these tests

15. Compared to regular class discussions,
   a. much harder on these tests
   b. a little harder on these tests
   c. about the same on these tests
   d. a little less hard on these tests
   e. much less hard on these tests

16. The essay directions asked you to imagine you were in the same historical period as the authors of the reading passages and to explain the meaning of events to another person. How did that part of the directions influence your performance on the essay?
   a. it helped me organize and choose information in my writing
   b. it neither helped or hurt
   c. it interfered with my writing
   d. it made it fun
   e. I didn't pay any attention to it at all

17. Which of these statements is true for you? Check all that apply
   a. I didn't know the information the test asked for
   b. The essay asked me to write in a way that I haven't been taught to do
   c. These tests were more fun than regular tests
   d. These tests were about the right level of difficulty

18. No matter how I did, I feel like I know the topic tested
   a. very well
   b. pretty well
   c. only a little
   d. not at all

19. How do you rate yourself as a writer?
   a. I am a very good writer
   b. I am a pretty good writer
   c. I am a fair writer
   d. I am not at all good at writing

20. How good do you think you are in history?
   a. I am very good in history
   b. I am pretty good in history
   c. I am fair in history
   d. I am not at all good in history

21. How good a student are you?
   a. I am a very good student
   b. I am a pretty good student
   c. I am a fair student
   d. I am not a good student
22. How do you think you could improve your performance on the essay task?
   a. much more time
   b. much more instruction on how to do it
   c. more practice
   d. I could try harder
   e. given opportunity to review some history content I have just learned
   f. no way I can improve
   g. some other way (please specify)

23. How do you think you could improve your performance on the semantic mapping task?
   a. much more time
   b. much more instruction on how to do it
   c. more practice
   d. I could try harder
   e. given opportunity to review some history content I have just learned
   f. no way I can improve
   g. some other way (please specify)
24. **Circle the appropriate number**

<table>
<thead>
<tr>
<th>How often do you:</th>
<th>At least once a week</th>
<th>A few times a week</th>
<th>once a month</th>
<th>less than once a month</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Take tests in history readings</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>B. Write in-class essays in history</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>C. Have homework in history</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>D. Write in-class essays in other classes</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>E. Write history essays at home</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>F. Do longer research papers in history</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>G. Take short answer tests in history</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>H. Take multiple-choice tests in history</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>
3. Student Packet: Chemistry/Soda Task

Prior Knowledge Test
Text
Essay Prompt
How Much Do You Know About Chemistry?

Directions: This is a list of terms related to high school chemistry. In the space after each term, write down what comes to mind drawing upon your knowledge of chemistry. A brief definition would be acceptable, or a brief explanation of why that law, principle, concept, or procedure is important in explaining chemical phenomena. If a term is general, give both a general definition and a specific example to show your understanding, if you can.

*Good Example: Periodic table:* An arrangement of chemical elements based on the order of their atomic numbers. Shows variation in most of their properties. Shows a natural division of elements into metals and nonmetals, inert gases, atomic weights.

Do not define the term by simply restating the same words.

*Bad Example: Electron level:* The level of the electron.

Even if you are not sure about your answer, but think you know something, feel free to guess. There are probably more items here than you will be able to answer in the time given. Start with the ones you know best, and work quickly so that you can answer as many as possible. Then go back and answer the ones of which you are less sure. Do not spend too much time on one specific item.

1. density

2. solubility test

3. conductivity

4. chemical reaction

5. base

6. nucleus

7. deductive reasoning
8. conservation of energy

9. precipitation

10. fructose

11. hypothesize

12. empirical formula

13. acid

14. experimental control

15. gas laws

16. compound

17. ion

18. indicator

19. quantitative analysis

20. hydration
Chemistry Demonstration: Soda Task

As an introduction to chemical analysis, a high school chemistry teacher performed an experiment for her class. This is a description of what she did:

"I have two samples of soda," she told the class, "one is regular soda containing sugar and the other is diet soda which contains an artificial sweetener. I'm going to identify each sample as diet or regular by doing some chemical tests. As in any chemical testing, I won't allow myself to taste the samples but will base my decision solely on the chemical and physical properties of the two samples as determined by the tests."

She began by labeling the samples A and B to help her keep track of the sample she was testing. She then proceeded by saying, "Since we've been studying the properties of many different kinds of substances, we know that we often can identify an unknown substance by performing physical and chemical tests on the substance and observing reactions. For example, acids turn certain solutions pink, while alkalis turn them green, and neutral ingredients fail to change the color of the solution. Keeping in mind the chemical properties of sugar, I'm going to conduct the following tests: the yeast test, the benedict solution test, a test using sulfuric acid, a solubility test, a test using salt, and a residue test."

Her first test was the yeast test. She poured equal amounts of each soda into separate test tubes and labeled them A and B respectively. One soda reacted with the yeast to give off a distinctive odor as well as gas bubbles, and the other did not react in the same way.

Next she used a benedict solution test. She began by pouring the indicator (benedict solution) into three test tubes. Into each container she added one of the two sodas, making sure to note on each test tube which one of the sodas was added. The third container was a control: nothing was added to the indicator in this test tube. She waited, knowing that some substances take a while to react with the indicator. Comparing the two tests tubes with the control, she pointed out that a reddish precipitate had formed in one of the test tubes.

For her next test, she mixed sulfuric acid with each of the sodas, handling the acid with extreme caution. She began by heating the sodas so that most of the liquid evaporated. Then as she added the sulfuric acid to each sample, she noticed that the acid reacted with one of the sodas to form a gooey brown substance.

To conduct the solubility test, she poured 100 ml of soda into separate beakers and gradually added equal amounts of sugar to each soda. She stirred the sodas and waited 15 seconds to see if the sugar dissolved. She found that more sugar dissolved in one soda than the other.

Next she added salt to each sample of soda. She noticed that as salt was added, one soda fizzed more than the other.

Finally, for the residue test, she placed 30 ml of each solution in separate test tubes, placed both tubes over a Bunsen burner and heated them until 15 ml evaporated from each. She noticed that more residue was left in one of the test tubes.

Upon completing the various tests, she made a chart of the results which looked like this:
| Yeast test     | distinct odor      | no odor     |
|               | gas bubbles       | no bubbles |
| Benedict solution test | reddish precipitate | no precipitate |
| Sulfuric acid test | produced a gooey brown substance | no gooey brown substance |
| Solubility test | not much sugar dissolved | a lot of sugar |
| Salt test      | not much fizzing   | a lot of fizzing |
| Residue test   | a lot of residue   | not much residue |

The teacher ended her demonstration by saying, "With your knowledge of the properties of sugar and the results of the tests, you now can determine which of these sodas is the regular and which is the diet."
Writing Assignment

Imagine you are taking a chemistry class with a teacher who has just given the demonstration of chemical analysis you read about earlier.

Since the start of the year, your class has been studying the principles and procedures used in chemical analysis. One of your friends has missed several weeks of class because of illness and is worried about a major exam in chemistry that will be given in two weeks. This friend asks you to explain everything that she will need to know for the exam.

Write an essay in which you explain the most important ideas and principles that your friend should understand. In your essay you should include general concepts and specific facts you know about chemistry, and especially what you know about chemical analysis or identifying unknown substances. You should also explain how the teacher's demonstration illustrates important principles of chemistry.

Be sure to show the relationships among the ideas, facts, and procedures you know.
PART V

SPECIFICATIONS FOR DEVELOPING ASSESSMENT MATERIALS
PART V:
SPECIFICATIONS FOR DEVELOPING ASSESSMENT MATERIALS

The demand for alternative ways to assess students' higher order thinking skills rather than mere recall of facts has resulted in a dramatic increase in the development of new measures by teachers and other experts for use at the local as well as state level. When many new performance tasks are developed independently by different authors, the coherence of the tasks is of particular concern. That is, do supposedly similar assessment tasks tap the same intellectual processes?

Comparability among tasks really doesn't matter in individual classrooms. Teachers have the obligation to use their own creativity and values in creating tasks that faithfully assess students in terms of their instructional experiences.

But as soon as we move into the accountability realm, the concern for task comparability increases significantly. Whether we are looking at individual accountability, such as the certification of students at the end of high school, program evaluation, or state assessment, we are fundamentally interested in making comparisons. When one makes comparisons, common measures must be used so that people can be compared fairly. And thus, critical elements of assessment tasks cannot vary at the will or whim of a teacher or school. A common template or framework must guide the development of such tasks. Such a template attempts to assure that assessment tasks share common features. Such a template is a set of assessment specifications.

Specifying Assessments

Specifications are explicit constraint statements that provide rules for the inclusion or exclusion of material in an assessment. The purpose of specifications is to permit the development of multiple, parallel assessment tasks that might be reasonably expected to assess students' subject matter understanding. The function of specifications is to control the behavior of the "item writer" or assessment designer and to provide cues about desirable and acceptable content and structure. Obviously, in an area as rich and complex as American history, even stringent specifications will restrict only to a degree the range and focus of any assessment. The trick is to control the critical features of the assessment.

Our CRESST project used specifications to control the structure and content of assessment tasks. In fact, the development of specifications for the essay task consumed a good deal of attention, particularly the creation of the scoring scheme, as described in Part II of this handbook.

The specifications below were designed to generate new tasks comparable to those used for the materials in this handbook. Specifications are provided here for the three major components of our assessment method, as noted in Table 1. If assessment requirements differ, obviously the specifications should be modified to reflect important aspects of local context or administrative constraints.

<table>
<thead>
<tr>
<th>Table 1. Assessment Components Requiring Specification</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Text Materials</td>
</tr>
<tr>
<td>2. Prior Knowledge Measure</td>
</tr>
<tr>
<td>3. Essay Task</td>
</tr>
</tbody>
</table>
Specifications for History Text Materials
The following set of specifications, enumerated in Table 2, was developed to cue the selection of text for use in assessing history.

Table 2
Specifications for History Text Materials

<table>
<thead>
<tr>
<th>Topic:</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>• Must be a regular and significant part of the secondary school</td>
</tr>
<tr>
<td></td>
<td>History Curriculum</td>
</tr>
<tr>
<td></td>
<td>• Must provide an issue that has implications beyond the</td>
</tr>
<tr>
<td></td>
<td>particular historical period</td>
</tr>
<tr>
<td>Text Structure and Form</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Must provide for contrasting views, explanations or contexts</td>
</tr>
<tr>
<td></td>
<td>• May use either a single piece or short contrasting pieces</td>
</tr>
<tr>
<td></td>
<td>• May be written in narrative or expository form</td>
</tr>
<tr>
<td></td>
<td>• Should be short enough to be read within a class period or</td>
</tr>
<tr>
<td></td>
<td>less</td>
</tr>
<tr>
<td></td>
<td>• Must be written so that esoteric or technical discussions are</td>
</tr>
<tr>
<td></td>
<td>minimal, special vocabulary is limited, and the author's point</td>
</tr>
<tr>
<td></td>
<td>of view is clear</td>
</tr>
<tr>
<td>Text Source</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Must use primary source materials.</td>
</tr>
<tr>
<td></td>
<td>• Acceptance pieces include letters, transcripts of speeches,</td>
</tr>
<tr>
<td></td>
<td>editorials, excerpts from documents.</td>
</tr>
<tr>
<td></td>
<td>• As a rule, provide only two major text selections. However,</td>
</tr>
<tr>
<td></td>
<td>supplementary materials, including written personal reactions by</td>
</tr>
<tr>
<td></td>
<td>historical figures, maps, songs, and other relevant material may</td>
</tr>
<tr>
<td></td>
<td>be made available, so long as adequate time is provided for the</td>
</tr>
<tr>
<td></td>
<td>students to encounter the materials. These additional materials</td>
</tr>
<tr>
<td></td>
<td>can be described as required or optional to the students.</td>
</tr>
<tr>
<td></td>
<td>• Editing of materials is to be avoided except to excerpt sections</td>
</tr>
<tr>
<td></td>
<td>from a longer piece.</td>
</tr>
</tbody>
</table>
Specifications for the Prior Knowledge Measure
The strategy for the measurement of prior knowledge involved the use of a 20 item, short answer assessment. The purposes for this measure were to: activate relevant prior knowledge for subsequent application in the essay; measure students' relevant prior knowledge in the subject matter; and get a general assessment of students' knowledge of American history. Because the forms of prior knowledge can range from broad principles to specific facts, our specifications reflect these forms as well as our multiple purposes. The specifications are included in Table 3.

Table 3
Specifications for the Prior Knowledge Measure

<table>
<thead>
<tr>
<th>Stimulus Format</th>
<th>* Proper names, terms, numbers, and short sentences not to exceed 8 words</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forms of Knowledge</td>
<td>* Specific facts, events, dates, quotations, the names of principles, or concepts.</td>
</tr>
<tr>
<td>Distribution</td>
<td>* Half the items should be reference specific information, such as an event (Harper's Ferry Incident) and half of the items should reference concepts (States' Rights) or principles (constitutionality).</td>
</tr>
<tr>
<td></td>
<td>* At least 2/3 of the items should be relevant to the immediate historic period of the assessment (+ or - 10 years of the date(s) of the text. The remaining 1/3 can precede or follow the period under assessment.</td>
</tr>
<tr>
<td>Directions to Students</td>
<td>* Students should be encouraged to respond rapidly, and to write the essence of their understanding briefly. No requirements for form, i.e., complete sentences, are required.</td>
</tr>
<tr>
<td>Administration Constraints</td>
<td>* Approximately 20 items can be provided in a 10-15 minute period.</td>
</tr>
<tr>
<td>Scoring Scheme</td>
<td>* Responses are scored on a five point scale (4=high). Students are given a 4 if they have an accurate, elaborated definition, description or context for the stimulus term; a &quot;3&quot; if they are essentially but minimally correct, a &quot;2&quot; is assigned if they have some incomplete notion of the term; a &quot;1&quot; if they have no idea; and a &quot;0&quot; if they make no response.</td>
</tr>
</tbody>
</table>

Specifications for The Essay Task
Specifications for the essay tasks focus on providing students sufficient cues for the type of answer desired, the form of the answer and the context for writing including a description of the intended audience. These issues are encompassed in the essay task specifications in Table 4.
| Essay Context | • Students should be given an historical context to frame the written response, consisting at minimum of a time, historical period, and occupational role. |
| Audience | • The audience for the student's writing is specified to be a particular person in the same target historical period. The purpose must be ignorant of the information provided in the texts for some plausible reason, i.e., living abroad, on a long trip, to heighten the verisimilitude of the task. |
| Intellectual Task | • The student needs to prepare an explanation of the dispute or topic included in the text selection(s). This explanation requires the student to understand the viewpoints expressed, compare and contrast perspectives using inference strategies, and to synthesize the explanation referring both to relevant text material and prior knowledge. |
| Directions to the Student | • Students should be given the directions that state the context and audience, and cue them related to critical format issues. The directions must underscore the need to use knowledge the student has acquired about history outside the text as well as to base the essay on the provided texts. |
| Administration constraints | • Directions can be printed at the top of sufficient paper for writing provided to the students. Students can also construct their answers using word processing equipment. |
| Scoring Scheme | • Students should have the text selections available to them as they write. |
| | • Students may be asked to complete the task in one class period (approximately 45 minutes), or they may be given a chance to revise their work. In the latter case, students should turn in their work at the close of each period. |
| | Essays should be scored in terms of the five scoring dimensions, Overall Content Quality, Principles, Prior Knowledge, Text, and Misconceptions, as described at length in the training materials. |
Following the use of specifications to generate comparable assessment tasks for periods of interest, i.e., the Westward expansion, the resulting draft assessment tasks should be reviewed independently by at least one knowledgeable person. This review entails comparing the specifications and tasks to assure that the tasks conform to the particular constraints.
PART VI
RATER TRAINING, SCORING, AND REPORTING PROCEDURES
PART VI: RATER TRAINING, SCORING, AND REPORTING

This section provides guidelines for organizing and conducting rater training and scoring sessions using the CRESST essay scoring rubric and prior knowledge scoring scale. It also addresses reporting options. By far the greater emphasis is placed here on the essay rating procedure because this is far more important and difficult than scoring the prior knowledge measure. The following procedures replicate those used at CRESST to conduct research on content assessment for both program evaluation and assessment of individual students.

Who Can Rate?

At CRESST, we have successfully trained raters with widely varying backgrounds, including practicing secondary school teachers, graduate students, subject-matter experts, elementary teachers, substitute teachers, and district office administrators. As essay scoring in a content area such as U.S. history is very different from the kinds of formal scoring experiences most educators have participated in, you will need to remind "experienced" raters that the task at hand may not resemble rating sessions they have participated in previously. We have found that people with strong content area backgrounds (e.g., U.S. history), either as an undergraduate or through later course work, make particularly good raters.

How Many Raters Do You Need?

Determining the optimal number of raters to recruit requires a little simple arithmetic. In our research, we've found that one reader can score about 75 essays a day before losing attention and precision. We have also found that when possible, it's better to use more raters for a shorter period of time to maintain a certain "momentum" to the rating. If you follow the 75-paper-per-reader-per-day guideline, all you need to do to determine the number of rater-days necessary is to divide the total number of papers you will be scoring by 75. When you have arrived at the number of rater days, simply divide that number by the number of days you have available for scoring and you will find the actual number of raters needed. Or, if you are constrained by the number of raters you can hire at any one time, divide the rater-days by the number of raters, and you will find out how long it will take to score your particular set of papers. The example below shows the process for determining how many raters will be needed.

**Calculation of Number of Raters Needed**

Example: You have 3,000 papers to score.
Guideline: One person can score about 75 papers per day (6 hour day).
Number of Rater Days: 3000/75=40 rater days
Guideline: More raters over fewer days, if possible.
Number of raters needed: 10 raters for 4 days or 8 raters for 5 days;

Note that the total number of essays to be scored should also include those that will be scored a second time and check papers.

When possible, train slightly more raters than you think you'll need. There may be attrition due to illness or you may find that some raters cannot be brought up to acceptable levels of reliability. If you have trained a few extra raters, then you have the option of tactfully dismissing the "aberrant" rater or assigning the person to a different task.

How Should the Training Be Organized?

Schedule the scoring session(s) well in advance so that you can recruit good raters. Training raters to use the CRESST essay scoring rubric takes three to four hours, depending upon how quickly the raters reach agreement on training papers. Training will take slightly longer if a rater is to score papers on several different topics (e.g. the Revolutionary War and the Civil War) because he or she will need time to become familiar with the primary text materials for each prompt and to practice applying each scale to writing on different topics.
You will need to schedule scoring so that you have the requisite 3-4 hours training time (or more) on the first day and perhaps 20-30 minutes of "refresher" training at the beginning of each new day. If your budget allows, try to limit actual training and rating time to 6 hours not including breaks and lunch. Schedule training with a brief informal period of 10 minutes or so at the start in which raters can meet and late arrivals will be less disruptive.

Sample Schedule for Training and Scoring Sessions

Day 1

8:00-8:10  Coffee and introductions
8:10-8:30  Introduction of the writing prompt, rubric
8:30-9:15  Review each scale in turn, with model essays to illustrate each scale; discussion of model essays
9:15-9:30  Break
9:30-10:00 Continue learning each scale with model essays
10:00-11:30 Raters practice applying whole rubric to training papers
11:30-12:15 Lunch
12:15-12:30 Complete training and reliability check with criterion papers
12:30-1:30 Score first set of papers (approximately 15)
1:30-2:30  Score second set of papers
2:30-2:45  Break
2:45-3:35  Score third set of papers
3:35-4:00  Debriefing, discussion of problems

Day 2

8:00-8:10  Coffee and assignment of raters to tables
8:10-8:30  Refresher training
8:30-9:30  Scoring of first set of papers
9:30-10:30 Scoring of second set of papers
10:30-10:45 Break
10:45-11:45 Scoring of third set of papers
11:45-1:00 Lunch
1:00-1:15  Refresher training (if needed)
1:15-2:15  Scoring of fourth set of papers
2:15-3:15  Scoring of fifth set of papers
3:15-3:30  Break
3:30-4:00  Debriefing and troubleshooting (if needed)

What Are the Trainer's Responsibilities?

The trainer should have considerable experience with standardized essay rating practices, either by having participated in scoring sessions or by having worked as a table or training session leader. The trainer has the following responsibilities:

1) To understand the scoring rubrics and be able to provide clear explanations of each of the score points,

2) To prepare an introduction to the training session by adapting the "script" provided with the CRESST materials to local conditions,

3) To set up the training and scoring schedule and select table leaders,
4) To read the model essays, training essays, criterion essays, and check essays and be able to explain how scores match the rubrics,

5) To organize the papers to be scored, have papers "coded" if necessary, organize into "sets" for raters, and bring to the scoring site,

6) To organize the reliability check at the end of training and schedule the check papers to maintain high rater reliability,

7) To retrain, dismiss or reassign "unreliable" raters,

8) To supervise the reproduction of training materials (prompts, rubrics, training papers, model essays, essay scoring summaries) and organization into rater packets, and

9) To devise the "set-up" for the training session, assignment of raters to different scoring "tables," and facilitate luncheon arrangements. Trainers may contact CRESST for assistance with the rubric if necessary.

**How Should the Room Be Set Up for Training and Scoring?**

If possible, select a quiet temperature-controlled room with space for several tables to seat raters comfortably (4-8 per table including the group leader). Ideally, an overhead projector will be available to project sample papers during rater training. Set aside space for storage of papers to be scored and already-scored papers, for conducting the reliability study, and for serving coffee, water, soft drinks and light refreshments during the day. In addition to the training and scoring materials, you should provide extra pencils and scratch paper. Access to a copy machine will enable you to discuss unanticipated problems arising from any papers that differ dramatically from those in the training materials.

**What Materials Need to Be Prepared for Training?**

The trainer and assistants will prepare materials for the training session. Examples of all materials necessary for rater training are contained in this handbook. The table below provides an overview of these training materials:

<table>
<thead>
<tr>
<th>Materials</th>
<th>Users</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Raters</td>
</tr>
<tr>
<td>1. Scoring Guidelines</td>
<td>X</td>
</tr>
<tr>
<td>2. Scoring Recording Forms</td>
<td>X</td>
</tr>
<tr>
<td>3. Prescored Typed Model Essays</td>
<td>X</td>
</tr>
<tr>
<td>4. Unscored Typed Model Essays</td>
<td>X</td>
</tr>
<tr>
<td>5. Prescored Training Essays - handwritten</td>
<td></td>
</tr>
<tr>
<td>6. Unscored Training Essays - handwritten</td>
<td>X</td>
</tr>
<tr>
<td>7. Prescored Criterion Essays</td>
<td></td>
</tr>
<tr>
<td>8. Unscored Criterion Essays</td>
<td>X</td>
</tr>
</tbody>
</table>
Allow plenty of time for reproduction of training materials and assembly into packets for raters. Raters may find it easier to have their materials organized into notebooks with appropriate dividers rather than an envelope or folder with loose papers. A suggested organization for the rater training notebooks appears below:

Suggested Organization of Rater Training Materials
(notebook with dividers)

Section I: Copy of prior knowledge test, primary text materials given students, and actual writing prompt. (If you are scoring papers for several prompts, keep these in separate sections.)
Section II: Copy of the short and long forms of the rubrics
Section III: Score Recording Forms
Section IV: Prescored and Unscored Model Essays to illustrate each scale (typed)
Section V: Prescored and Unscored Training Essays (handwritten)
Section VI: Criterion Essays (handwritten)
(repeat for each separate topic)

It is estimated that a minimum of twenty papers needs to be available for training in a single task and approximately three to four more for each additional topic area. Part VII contains sets of such papers for training purposes. These papers are for duplication. The conceptual background paper in Part II may be shared with raters in advance of the training sessions to provide some common ground.

What Training Procedures Should Be Followed?

Introductions. The trainer should start the session by introducing himself or herself and have raters introduce themselves by describing what they teach, who their students are, any previous rating experience, and any concerns they have about the upcoming process. As there are several "ice breaking" techniques to help raters get to know one another, the trainer can select any method that would be appropriate for the group and the time available. Once people have been introduced, the trainer may begin by explaining the purpose of the session, the goals of training, and the importance of training to assure consistent scoring of papers.

Familiarization with the writing tasks. The trainer may then ask the raters to turn to the first section of the scoring notebooks and read the writing task presented to students (the primary source materials and the actual prompt). If necessary, the trainer will answer questions about the tasks and clarify the writing situation.

Introduction to the scoring rubric and individual scales. Raters are then directed to read the scoring rubric, and the trainer will provide an overview of the five scales. When questions about the rubric as a whole have been answered, training should move on to the specifics of each scale, using the typed, prescored Model Essays to illustrate each scale. During training it may be helpful for both trainer and trainees to make note of the factual information and ideas contained in the primary source text so as to recognize them later when using the Proportion of Text Detail scale. In addition, raters may find it useful to list some of the concepts or principles that may frequently occur in the essays, to facilitate initial application of the Number of Principles scale.

Eventually the trainer can move to an unscored essay to let the raters attempt to apply the scale in question for themselves, referring to the scored version for feedback. The trainer should not move on to explain a new scale until the group understands the current scale well. After all five scales have been presented and learned, the trainer should provide raters with practice using all the scales on a single essay.

Training with the full rubric. Raters will be asked to apply the entire rubric to an unscored version of one of the typed model essays (if any are left unused at this point) or to a
handwritten training essay, after which the trainer will provide corrective feedback using the prescored versions.

In applying the full rubric, the rater is asked to read the paper and always score first the Overall Impression of Content Quality scale, and then to go through the paper carefully to complete the scoring on the other scales. If questioned about this sequence, the trainer should point out that if the raters were to score an overall category after scoring the other scales, this holistic score would lose some of its validity. Our research shows that when overall judgments follow rather than precede subscores, they may become simple summaries or averages of scores given to other more focused parts of the scoring rubric. The trainer should provide feedback on the correct scores for each training paper, scale by scale. The trainer should ask each rater to report his/her score or in the case of a larger group, to show by raising hands what score points were assigned to the paper under review. The trainer should always note the portions of the essay that support the valid score point.

The trainer should eventually move to the essays provided in student handwriting. This sequence, from typed to student scripted essays, increases the similarity to actual scoring. If student papers to be scored in the real scoring session will be provided in typed form, such as those completed on a word processor, then the student scripted segment could be skipped.

Rater reliability assessment. The last training task is to assure that raters:
1) are assigning scores on the basis of the rubric,
2) agree with each other in the assignment of scores, and
3) rate at a reasonable speed (about 3-5 minutes per essay).

In order to assess rater reliability (that is, fidelity to the rubric and consistency with each other), you will need to run a small reliability assessment using the "criterion" papers. The criterion papers are not model essays, training essays, or practice essays. They are a special set of 3-5 essays to be used at the end of the training period. At this times, the trainer should give each rater the same packet of criterion papers and tell them that this packet is a trial run for the scoring session and will be used to assess the effectiveness of training. Thus, raters should use their rubric guidelines and score each paper in the packet as though it were under actual scoring conditions. Raters are not to talk to each other during this reliability assessment, nor will there be a discussion of the papers afterward. Once the papers are scored and the scores are recorded on the Scoring Sheet, raters should be dismissed for a break or lunch while the trainer and an assistant conducts the reliability check.

To calculate rater reliability on the criterion papers, follow these steps and see the example in Table 6:

1) Have ready the Criterion Papers Scoring Sheet for Reliability Check (in Part VII) with highlighted squares for the "correct" rating for each scale.

2) Record the number of raters (frequency count) who selected each scale point. In the example, for the first criterion paper, 5 raters gave the paper a "2" on the GICQ scale, which was the "correct" rating.

3) Count the number of raters who gave the paper the correct rating. These are the raters with "perfect agreement." Enter the number in the appropriate column.

4) Calculate the percent of the total number of raters who have "perfect agreement." In the example, this is 50% for Paper #1. Record this number on your chart.

5) Count the number of raters who gave the paper a rating that is within 1 point higher or lower than the correct rating. These raters are in the "criterion score ±1 agreement" range. Record this number on your chart.

6) Calculate the percent of the total number of raters who are in the ±1 agreement range. In the example, this is 90% for Paper #1. Record this number on your chart.
7) Repeat for each scale, for each criterion paper.

8) Copy the "Percent Perfect Agreement" and the "Percent ±1 Agreement" in the "Averaging Table" at the bottom of the chart.

9) Calculate the average for each of the two kinds of percent agreement for each scale. In the example, the average degree of "perfect agreement" for the GICQ scale for the 10 raters who scored Criterion Papers #1 and #2 is only 65%, but the average "±1 Agreement" for that scale is 90%.
Table 6

Calculating Reliability Based on Criterion Papers

<table>
<thead>
<tr>
<th>Paper #1 Scales</th>
<th>Frequency Count of Scores</th>
<th># of criterion ratings</th>
<th>% perfect agreement</th>
<th># of C±1 ratings</th>
<th>%±1 agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td>GICQ</td>
<td>11 1 1 1</td>
<td>5</td>
<td>50%</td>
<td>9</td>
<td>90%</td>
</tr>
<tr>
<td>NP</td>
<td>1 1 11 1</td>
<td>7</td>
<td>70%</td>
<td>10</td>
<td>100%</td>
</tr>
<tr>
<td>PK</td>
<td>1 1 1 1 1</td>
<td>6</td>
<td>60%</td>
<td>8</td>
<td>80%</td>
</tr>
<tr>
<td>TEXT</td>
<td>1 1 1 1 1 1 1</td>
<td>10</td>
<td>100%</td>
<td>10</td>
<td>100%</td>
</tr>
<tr>
<td>MIS</td>
<td>1 1 1 1 1 1 1</td>
<td>6</td>
<td>60%</td>
<td>8</td>
<td>80%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Paper #2 Scales</th>
<th>Frequency Count of Scores</th>
<th># of criterion ratings</th>
<th>% perfect agreement</th>
<th># of C±1 ratings</th>
<th>%±1 agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td>GICQ</td>
<td>1 1 1 1 1</td>
<td>8</td>
<td>80%</td>
<td>9</td>
<td>90%</td>
</tr>
<tr>
<td>NP</td>
<td>1 1 1 1 1 1</td>
<td>8</td>
<td>80%</td>
<td>10</td>
<td>100%</td>
</tr>
<tr>
<td>PK</td>
<td>1 1 1 1 1</td>
<td>9</td>
<td>90%</td>
<td>9</td>
<td>90%</td>
</tr>
<tr>
<td>TEXT</td>
<td>1 1 1 1 1 1</td>
<td>8</td>
<td>80%</td>
<td>10</td>
<td>100%</td>
</tr>
<tr>
<td>MIS</td>
<td>1 1 1 1 1</td>
<td>7</td>
<td>70%</td>
<td>9</td>
<td>90%</td>
</tr>
</tbody>
</table>

Total # of Raters = 10

Averaging Table

<table>
<thead>
<tr>
<th>Scales</th>
<th>Perfect Agreement</th>
<th>±1 Agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Paper #1</td>
<td>Paper #2</td>
</tr>
<tr>
<td>GICQ</td>
<td>50%</td>
<td>80%</td>
</tr>
<tr>
<td>NP</td>
<td>70%</td>
<td>80%</td>
</tr>
<tr>
<td>PK</td>
<td>60%</td>
<td>90%</td>
</tr>
<tr>
<td>TEXT</td>
<td>100</td>
<td>80%</td>
</tr>
<tr>
<td>MIS</td>
<td>60%</td>
<td>70%</td>
</tr>
</tbody>
</table>
The criterion levels for rater agreement vary with the purpose of the task, with high stakes decisions for individuals requiring the highest level of agreement. In general, Percent Perfect Agreement should be fairly high (at least 80%), and the Plus-or-Minus-One Agreement should be even higher (e.g. 90-100%) because the latter is actually quite a liberal difference from the criterion score on a scale with only six points. The trainer should attempt to train all raters to obtain the criterion score for each scale, not merely within a point in either direction.

If a small group of raters are inconsistent with the rest of the group, you may wish to retrain this small group so that you can have the benefit of a larger number of raters for the actual scoring. If only one or two raters are assigning scores inconsistently with the group, the following options are available:

1. Dismiss the rater in a tactful way from the rating task. Give the rater something else to do, since he or she is probably being compensated for time. A specific recommendation is to score the prior knowledge tests for the same tasks, directions for which follow. It is our experience that scoring the prior knowledge task is a much simpler problem.

2. Retrain the rater. This option is one which is particularly desirable if only a few raters are available, if time is constrained, and if the divergent rater has problems on only one or two scoring dimensions. This retraining can occur after starting the entire group on the rating of the target essays.

   If low scores would result in significant consequences for student, the trainer should also be on the lookout for raters who consistently tend to rate inaccurately toward the low side, for their bias may negatively affect important decisions. Additional training for such raters may be necessary.

   If a large proportion of the raters is not reaching agreement, you will need to retrain and conduct another reliability assessment before proceeding to the scoring session. If this does not succeed, the following options are available:

1. Decide to go with a longer rating period and many fewer raters.

2. Decide to discard an element of the scoring scheme on which agreement cannot be reached. This is a last resort.

3. Decide that the trainer/raters selected are not up to the job and find others (in practice only possible in research conditions where budget and time constraints have some flexibility). This is an act of desperation and should not be necessary.

**What Materials Need to Be Prepared for Scoring?**

Table 7 lists the materials you will need for the actual scoring session. Examples of these materials are included in Part VII.
Table 7
Materials Required for the Scoring of Local Essays

<table>
<thead>
<tr>
<th>Materials</th>
<th>Users</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Packets of Student Essays with Interpolated Check Papers</td>
<td>Raters</td>
</tr>
<tr>
<td>2. Prescored Check Papers</td>
<td></td>
</tr>
<tr>
<td>3. Scoring Guidelines</td>
<td>X</td>
</tr>
<tr>
<td>4. Scoring Sheets</td>
<td>X</td>
</tr>
<tr>
<td>5. Calculator</td>
<td></td>
</tr>
</tbody>
</table>

Papers will need to be prepared before giving them to raters and may be prepared well in advance of training and scoring.

**Single or multiple scores per paper.** Prior to scoring papers, you must decide whether each paper will be read one, two, or more times. While we commonly assume that multiple scores of a single essay will yield a more reliable measure of a student’s writing, the fact is the need for multiple scores depends upon the purpose of the assessment. If student scores will be used for program evaluation rather than individual diagnosis, and especially if the writing prompts were sampled (each student did not get the same prompt), a reliable estimate of an individual student score is less critical than a reliable estimate of the average score for the task. In this case, most papers may be read only once, with reliability assured for the set of papers by having only a **sample** of the papers scored more than once. You can then calculate rater agreement much as was done to assure reliability during rater training or using more sophisticated statistical techniques to assure that raters are assigning scores consistently with each other and in accordance with the rubric.

The sample of papers that receive multiple scores should represent approximately 20% of the entire paper set. They need to be reproduced and inserted in every rater’s packet in a different random order. The 20% sample should represent as much as possible the entire range of the population being assessed, which would include different ability levels, school sites, language backgrounds or other variables that may be related to program effects. If time and cost make it difficult for each rater to score the entire 20% sample of papers, you may assign papers to pairs of raters. To do this correctly, make sure that the rater pairs are randomly assigned to each paper and that every rater is paired with every other rater on at least one occasion. In other words, you don’t want to have the same two people scoring the entire 20% sample. All raters should rate some of the papers in this sample.

If scores will be used to make important decisions about **individual students**, multiple scoring of each paper is advised. Given even moderately successful training, however, no more than two raters will need to read any one paper. The two ratings can then be averaged or summed to provide the final score. No third rating is necessary to adjudicate scores. In situations where student scores will be used to make selection, placement, or instructional decisions for individuals based on some prespecified standard or “cut score”, you could provide a single rating for all papers and rescore only those falling within one point of either side of the cut score. If you are rescoring only those marginal papers, you will need to monitor scoring and have an assistant pick up papers and assign them to a second reader during the scoring session. It is important to direct raters not to mark scores, comments, or corrections on any papers that will receive multiple scores so as not to influence the second reading.
Organizing set of papers. Once you have developed your rating plan (e.g., single scores with a 20% multiply-scored sample, double scores on marginal papers, or multiple scores for all) you will need to assign raters their initial sets of papers and devise a system for getting double-scored papers to raters if they have not been integrated into their packets in advance. Next, distribute a pre-sorted packet of papers to each rater.

Papers should be randomly ordered within each rating packet and randomly assigned to raters. Advance preparation includes:

1) Assign and mark student identification numbers on each paper if this has not already been done. Commonly, districts use the student's social security number or district identification number.

2) If more than one writing prompt will be scored, either because each student wrote on more than one topic or because two different topics were used, a code number for "topic" or writing task should be assigned to each paper.

3) Sort papers by topic so that all papers related to one prompt are together and separated from papers on a different prompt.

4) If different grade levels have written on the same prompt, sort papers again by grade level.

5) Once papers are organized by writing prompt and grade level, randomly order the papers for each prompt, keeping papers on different prompts separate.

6) Divide the papers into groups by dividing the number of papers for each prompt by the number of raters. For example, if there are 1000 papers on Prompt A and 10 raters, create 10 groups of 100 papers each for Prompt A, then divide each rater's 100 papers into rating sets of 15.

7) If two or more prompts have been used, balance the order of prompts across raters on a given scoring day. For example, if there are 8 raters and 2 prompts, 4 raters should rate sets of Prompt A papers in the morning and Prompt B papers in the afternoon, while the other 4 raters should rate Prompt B in the morning and Prompt A in the afternoon. In our experience, raters found it difficult to shift from one topic to another with each new paper read. It was much easier for them to read one to two sets on one prompt and then shift to a new prompt.

8) Reproduce the "check" papers included with your CRESST materials so that you have one complete set for each rater. Insert the same check paper in the same position in each rater's packet. Typically, these check papers are inserted after each 12-15 papers. You will want to have at least 4 check papers read per day, perhaps as many as one per hour (raters typically read 10-15 essays an hour).

Scoring sheet. The second part of preparation for scoring is the creation of a Scoring Sheet for raters to enter information about student scores. Ideally, you will provide a format that lets raters enter each student's identification number and essay score on a machine-readable document so that scores may be quickly calculated and results published. We have included a sample Scoring Sheet in Part VII which you can adapt to meet your local situation.

What Scoring Procedures Should Be Followed?

Once your training has established acceptable reliability levels for your scoring purposes, you may begin scoring the actual essays. Be sure that raters have pencils, scratch paper, and scoring sheets upon which to record their scores, and instruct raters to begin.

During scoring. During scoring, the trainer and assistants should circulate to be sure that raters are entering relevant data on the scoring sheets, keeping their preassembled scoring packets together and scoring papers in the order in which they were placed in the packet.

Check papers. The use of multiple scores on a sample of papers, and common check papers inserted every nth paper provide a system for monitoring rater reliability, an especially
difficult task when the number of raters exceeds 20 or when rating is done at different sites and upon different occasions. The need to monitor rater fidelity to the rubric and consistency with each other occurs because raters tend to redefine the scoring system when they interact with their peers during training, scoring, and breaks. Double scored papers and check papers allow you to interrupt scoring, retrain, and prevent "rater drift" from the rubric. On the other hand, acceptable reliability is fundamentally dependent upon the quality of rater training. It is more economical to train raters well and not begin scoring until Interrater agreement reaches high levels than to retrain when rater agreement begins to deteriorate.

If sets of papers have been prepared in advance for raters as suggested previously, raters will encounter about one check paper an hour, or every two hours. At the minimum, we recommend at least one check paper in the morning and one in the afternoon. If you have inserted check papers to review each hour, at the end of an hour ask the rater to pull their scoring sheets and paper number XXX and collect the scores for these papers. Do not tell raters in advance which papers are check papers so as not to influence their "rating habits".

The trainer should calculate rater agreement using the same method used to calculate rater reliability with Criterion Papers during training. If raters are having widespread difficulty, the trainer can meet with the group in question; if there are only one or two errant raters, the trainer can work with these in a small group. If, over a series of check papers, a previously qualified rater begins to "lose it", the trainer can retrain the individual, reassign him or her to another task or dismiss the rater.

Reporting the Results of the Essay Task

One interesting way of using the results of the writing task assessment is to compare an individual or group's scores on the five scales to profiles of the typical expert or novice. As Figure 1 indicates, for example, a content expert tends to score high on General Impression of Content Quality, Number of Principles, and Prior Knowledge, but moderate on Text and low on Misconceptions. A content novice tends to do the opposite, to exhibit little conceptual or prior knowledge and to rely heavily on the text. Some novices have many misconceptions but others have few or are aware that their knowledge is shaky and manage to avoid revealing many misconceptions in their essays. Two students' work is depicted in Figure 2. Both students' essays had low content quality and high prior knowledge; however, the student who wrote Training essay #6 showed more conceptual knowledge and (perhaps as a consequence) more use of text detail than the student who wrote Check paper #2.
Figure 1
EXPERT AND NOVICE PROFILES

- expert
- novice
Figure 2
SAMPLE STUDENT PROFILES

check essay 2
training essay 6
Comparing an individual's or group's observed profile to expert and novice profiles can provide suggestions for individual diagnosis or program improvement. Of course both kinds of decisions are stronger if made on the basis of several measurements (or prompts) rather than one. A classroom teacher might use the results of several measures over time to depict student growth and strengths.

Comparisons of different groups' profiles or their gain scores over a year or more may also be made. It is helpful, however, to consider the entire distribution of scores, not simply the average. For example, you may be particularly interested in what happened to the students in the lowest scoring quartile over the course of a year. If statistical packages are readily available, analyses of covariance or regression analyses may be used to examine the effects of such variables as language ability or a special instructional program on performance.

Scoring and Using the Prior Knowledge Test

Scoring the Prior Knowledge Test follows much the same procedures as outlined above but is much quicker and simpler. Each of the 20 items on the test is given a score of 0, 1, 2, 3, or 4 based on the completeness and quality of students' responses. Here writing quality and grammar definitely do not count. Scoring guidelines and model tests may be found in Part VII for your use. In our experience, it takes about 1-2 minutes per test to score it, depending on the student's level of knowledge.

Generally, raters can be trained rather rapidly to high levels of agreement (95-100\% "perfect agreement") by seeing model answers of the highest level "4", assigning "3" when the student shows understanding, "2" when some element is correct, and "1" when the student misses the point. Zero is reserved for items the student did not even attempt to answer (blanks). Plus-or-minus-one-agreement may be skipped as too liberal for this four point scale. It is usually sufficient to double score only a 20\% sample of the Prior Knowledge Tests.

As a minimum, it is expected that the rater has thorough knowledge of the specific content area of interest (e.g. the Civil War era) and general content knowledge as well (e.g. U.S. history in general). Some raters who cannot seem to adapt to the essay scoring rubric can apply the prior knowledge rubric quite adequately.

If the results of this test are to be used in high stakes decisions, it may be necessary to create and use criterion tests and check tests, as in the essay scoring procedure, to ensure and document that appropriate procedures were followed.

The results of the Prior Knowledge Test can be reported and used in several ways, depending on your purpose. For example, individual item scores may be compared to provide information for improvement of local curriculum and instruction. For example, if very few students were able to define a particular term, teachers could examine their curriculum and instructional practices for an explanation and possible revision. If this approach is used at a classroom level, a teacher might note which prior knowledge items were utilized on the subsequent essay and which were not, even though they may have been known. Another way of using the findings to improve instruction is to group the items into subsets and use their total or average, such as a "factual knowledge" subscore and a "conceptual knowledge" subscore. Alternatively, it may be sufficient to have a total or average score for the entire test. In some circumstances, the test may be given more for its function to activate students' knowledge in order to improve their writing than to actually use the prior knowledge scores themselves. In this case, the test might not be scored at all or might be scored by the students themselves.
PART VII

SAMPLE TRAINING MATERIALS
PART VII: SAMPLE TRAINING MATERIALS

This section contains the following materials:
1. Prior knowledge scoring materials
2. Essay scoring guidelines and record forms
3. Student essays and scoring keys.
1. Prior Knowledge Scoring Materials

Scoring Guidelines
Scoring Sheet
Sample Student Responses
Prior Knowledge Scoring Guidelines

0= no answer or a completely wrong answer that indicates no understanding of the term.

1= A) a correct association but conveys no understanding of meaning or significance of term. For example, a name as the only response to any term.

EX: Dred Scott - slave
Communism - Russia

An association should not be counted if it is simply a reordering of the words of the term with no other words added.

EX: westward movement - moving west
industrialization - industry

B) an incorrect definition that suggests some correct association

EX: Kansas-Nebraska Act - one slave, one free (incorrect but identifies the act as related to the issue of slavery)
Dred Scott - freed slave who helped others escape

2= indication of partial understanding or some knowledge of the term.

EX: Missouri Compromise - Missouri enters as slave state
New Deal - Plan to help nation's economy
States' Rights - states can decide on things

3= indication of complete understanding or knowledge of the term (but does not have to be an elaborate definition)

EX: Underground Railroad - route by which slaves escaped from the south
Missouri Compromise - declared that no slavery north of 3630. Or Missouri entered as a slave state, Maine entered as a free state.

4= A) indication of complete understanding or knowledge of the term and its significance, cause or impact.

EX: Dred Scott - slave who sued for his freedom after being taken into free territory and lost. Ruled that slaves are property.
Industrialization - shift from agriculture to making things by machines. Contributes to growth of cities/source of difference between north and south.

B) an elaborated definition that indicates complete understanding or knowledge of term. For example adding a name, date or place to a score of 3.

EX: Gold Rush - 1849, gold discovered at Sutter's Mill in California, and many people went there to strike it rich.
General Rules

Answers should be upgraded one point if a correct detail such as a person, place, or date is added.

EX: Missouri Compromise - compromise between North and South = 1. If date is added = 2.

Errors should be ignored and the part of the answer that is correct should be scored. For example if the student identifies the New Deal as FDR’s program to help the nation recover from WWII, the answer should be scored as "FDR’s recovery program."
Prior Knowledge Scoring Sheet

<table>
<thead>
<tr>
<th>Student ID</th>
<th>Version</th>
<th>1-4</th>
<th>5-8</th>
<th>9-12</th>
<th>13-16</th>
<th>17-20</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Directions: This is a list of terms related to U.S. History. Most of them are related to the period of the Civil War, but some of them are from other periods in U.S. History.

In the space after each term, write down what comes to mind when you think of that term in the context of U.S. History. A brief definition would be acceptable, or a brief explanation of why that person, place, or thing was important. If the term is general, such as "Civil rights," give both a general definition and a specific example of how the term fits into U.S. History, if you can.

Good Example: Civil rights. Rights guaranteed to all citizens regardless of race, sex, religion, etc. Blacks fought for their civil rights in the 1960s. Martin Luther King, Montgomery bus boycott.

Do not define the term by simply restating the same words.

Bad Example: Survival of the fittest – only the fittest survive.

Even if you are not sure about your answer, but think you know something, feel free to guess.

There are probably more items here than you will be able to answer in the time given. Start with the ones you know best, and work quickly so that you can answer as many as possible. Then go back and answer the ones of which you are less sure. Do not spend too much time on one specific item.

1. popular sovereignty

2. Dred Scott
   A black slave who thought he should be free after his master took him into a free state. He lost in the Dred Scott decision.
3. Communism

4. Missouri Compromise
   all states below the southern part of Missouri were slave states
   and all above that (except Missouri) were free states

5. Industrialization

6. Gold Rush
   when someone discovered gold in an area where no one
   knew about. The California gold rush in 1849.

7. Bleeding Kansas
   the fighting that was in Kansas before the Civil War.

8. States' rights
   the laws a state has that are not the same as others

9. Federalism

10. Underground Railroad
    an escape for slaves

11. Imperialism
12. Whig Party that ran for president

13. Kansas-Nebraska Act

14. Abolitionists
   They were for the abolition of slavery

15. sectionalism
   When the country started moving apart.

16. westward movement
   Expanding the U.S. towards California.

17. constitutionality
   Having to do with the constitution.

18. New Deal

19. party platform
   The reason a person wants to be president is
   his party platform.

20. balance of power
   The legislative, executive, and judicial system all check
   each other to make sure they don't do anything wrong.

Stop time 1:35

TURN TO THE NEXT PAGE
Directions: This is a list of terms related to U.S. History. Most of them are related to the period of the Civil War, but some of them are from other periods in U.S. History.

In the space after each term, write down what comes to mind when you think of that term in the context of U.S. History. A brief definition would be acceptable, or a brief explanation of why that person, place, or thing was important. If the term is general, such as "Civil rights," give both a general definition and a specific example of how the term fits into U.S. History, if you can.

Good Example: Civil rights. Rights guaranteed to all citizens regardless of race, sex, religion, etc. Blacks fought for their civil rights in the 1960s. Martin Luther King, Montgomery bus boycott.

Do not define the term by simply restating the same words.

Bad Example: Survival of the fittest – only the fittest survive.

Even if you are not sure about your answer, but think you know something, feel free to guess.

There are probably more items here than you will be able to answer in the time given. Start with the ones you know best, and work quickly so that you can answer as many as possible. Then go back and answer the ones of which you are less sure. Do not spend too much time on one specific item.

1. popular sovereignty

2. Dred Scott Supreme Court decision on slavery.

Man's soul is in the whole man.
3. Communism
   A form of government that does not give people freedom. Ex. Russia

4. Missouri Compromise
   So that half of states would go
   so that there would be slavery / non-slavery

5. Industrialization
   The coming of machinery, factories, business
   19th century

6. Gold Rush
   In California: gold was found, thousands rushed to try to find it

7. Bleeding Kansas

8. States' rights
   The state is allowed to have laws
   so it can govern

9. Federalism

10. Underground railroad
    To help slaves get to the north

11. Imperialism
    Part of government
12. Whig party of the government district last long.

13. Kansas-Nebraska Act
A land act that said you could buy much land at minimum. Some much would by surplus for public use.

14. Abolitionists
They wanted slavery ended.

15. Sectionalism

16. Westward movement
People moved outward because land was cheap.

17. Constitutionality

18. New Deal

19. Party platform
Where a party can get up and speak its views

20. Balance of power
During WW2 this was disrupted because Italy became a country 2 years of the war.

Stop time 1:34. TURN TO THE NEXT PAGE
2. **Essay Scoring Materials**

- Essay Scoring Guidelines
- Essay Scoring Rubric
- Essay Rating Sheet
- Scanable Essay Scoring Sheet
Essay Scoring Guidelines

1. **General Impression - Content Quality (CICQ)**
   How well does the student know and understand this historical content?
   (0 - 5 point global rating: 0 = no response, 5 = highest level of understanding)

2. **Principles/Concepts - Number (PN)**
   This is a measure of the number of different social studies concepts or principles that the student uses with comprehension.

   A **concept** is an abstract, general notion, such as "inflation." It does not refer to particular events or objects (such as one particular period of inflation), but instead represents features common to a category or events or objects. "Imperialism," for example, does not refer to any specific facts or events; it is a heading that characterizes a class of behaviors and beliefs. "Industrialization" likewise identifies a class of activities and events that share common properties. It must be clear that the student is using a term conceptually, not just as a label.

   A **principle** is a rule or belief used to justify an action or judgment, as in the statement "Slavery is immoral," where "morality" serves as a justifying principle.

   It should be evident that the student understands the concept and means to discuss it. The concept should not simply be mentioned within a quotation from the text with no indication that the student grasps the concept.

   To earn a score point, the concept or principle need not be named explicitly, such as, "Constitutionality was an important principle that influenced the debate over slavery," but the idea should be stated clearly, for example, "One problem was determining what the constitution said about slavery."

   **Score point guidelines:**
   0 - no response
   1 - no concepts
   2 - one concept
   3 - two concepts
   4 - three concepts
   5 - four or more concepts

   **Adjustments:**

   Adjust up one score point if the principle(s) are referred to explicitly on a conceptual level: there is understanding of the significance of the concept.

   Example: "One great factor which held us back from war was our economy. It was not known what would happen to our economy without the safety of Britain. Britain could defend our commerce and coasts. Also, with Britain there was a great advantage with exportation. It seemed our economy could only suffer without the aid of Britain"

3. **Prior Knowledge: Facts and Events (PK)**
   This is a measure of the extent to which students incorporate relevant concrete information that is not mentioned in the speeches into their essay. This type of information may include pieces of legislation, court decisions, names of people, places or events, and general information about the period.
Statements of opinion are not included (e.g. "Lincoln was our greatest president"). Student should not be penalized for information that is incorrect; e.g. "In 1770 some people moved from England to America and started the first settlement there." (Mistakes will be accounted for in another scale.) Extremely common knowledge such as "slaves came from Africa" is not counted in this context.

Score point guidelines
0 - no response
1 - no facts/events mentioned that are not found in the
   texts of the speeches
2 - one to two facts/events
3 - three to four facts/events
4 - five to six facts/events
5 - seven or more facts/events

Adjustments

Adjust up one score point if the fact/event is very well described and the student demonstrates understanding of not only the meaning, but the significance.

Example: At Harper's Ferry John Brown attempted to lead a slave revolt but failed. This incident made the North and the South more divided because it convinced the Southerners that he abolitionists were crazy people who would do anything, even violence, to stop slavery.

4. Proportion of Text Detail (TEXT)
   This is a measure of the amount of material from the text of the speeches that is used in the essay.

   A text detail is a quotation, paraphrase, or any other reference to information and ideas in texts provided. It should be clear that the text detail in the student essay was extracted or learned from the texts provided. If you believe that the student did not obtain information from the texts provided, do not count it as a text detail.

   Assign score according to point which comes closest to the proportion of text detail in the student's essay. (e.g. In a long essay, 1 or 2 sentences reflecting information from the provided text will earn a TEXT score of 1.)

Score point guidelines
0 - no response
1 - no information from text
2 - material from the text accounts for about 1/4 of the essay
3 - material from text accounts for about 1/2 of the essay
4 - material from the text accounts for about 3/4 of the essay
5 - the essay uses or is based on material from the text only

5. Misconceptions (MIS)
   This is a measure of the amount of incorrect information, or the number of misconceptions or misinterpretations, in the essay. A higher score indicates more errors.

   Three possible types of errors to consider:
   - factual errors such as incorrect names or dates
   - misconceptions about the historical period
   - misunderstanding or misinterpretation of the text of the debates
Score point guidelines:
0 - no response
1 - no misconceptions
2 - very minor misconception
3 - several minor errors and/or a moderate misconception
4 - at least one serious misconception
5 - one or more serious misconceptions central to the essay

Wrong opinions or judgments (e.g., "Douglas made good arguments and Lincoln didn't know what he was taking about") are not counted as misconceptions.
Content Assessment
Scoring Rubric

1. **General Impression Content Quality (GICQ)**
   How well does the student know and understand this historical content?
   (0-5 point global rating: 0 = no response, 5 = highest level of understanding)

2. **Principles/Concepts - Number (PN)**
   0 - no response
   1 - no principles
   2 - one principle
   3 - two principles
   4 - three principles
   5 - four or more principles

   Adjust up one score point if the principle(s) are discussed explicitly on a conceptual level.

3. **Prior Knowledge: Facts and Events (PK)**
   0 - no response
   1 - no facts/events mentioned that are not found in the text of the debates
   2 - one to two pieces of information that are not found in the text of the debates
   3 - three to four pieces of information that are not found in the text of the debates
   4 - five to six pieces of information that are not found in the text of the debates
   5 - seven or more pieces of information that are not found in the text of the debates

   Adjust up one score point if the fact/event is very well described and the student demonstrates understanding of meaning and significance.

4. **Proportion of Text Detail (TEXT)**
   0 - no response
   1 - no information from text
   2 - material from the text accounts for about 1/4 of the essay
   3 - material from text accounts for about 1/2 of the essay
   4 - material from the text accounts for about 3/4 of the essay
   5 - the essay uses or is based on material from the text only

5. **Misconceptions (MIS)**
   0 - no response
   1 - no misconceptions
   2 - very minor
   3 - several minor errors and/or a moderate misconception
   4 - at least one serious misconception
   5 - one or more serious misconceptions central to the essay
<table>
<thead>
<tr>
<th>Student ID</th>
<th>Rater</th>
<th>Topic</th>
<th>1 GICQ</th>
<th>2 PN</th>
<th>3 PK</th>
<th>4 TEXT</th>
<th>5 MIS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
# Essay Scoring Sheet

**UCLA - Center for the Study of Evaluation**

<table>
<thead>
<tr>
<th>SITE</th>
<th>GROUP</th>
<th>CODE</th>
<th>STUDENT ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>MONTH</th>
<th>YEAR</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>7</td>
<td>7</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TASK</th>
<th>PROMPT</th>
<th>CONTENT</th>
<th>DRAFT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>X</td>
<td>X</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>X</td>
<td>X</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>X</td>
<td>X</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>X</td>
<td>X</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>X</td>
<td>X</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>X</td>
<td>X</td>
<td>1</td>
</tr>
</tbody>
</table>

## History Content Quality Scale

<table>
<thead>
<tr>
<th>RATER ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>7</td>
</tr>
</tbody>
</table>

### General Impression - Content Quality

### General Impression - Expression

### Principles - Number

### Principles - Function

### Facts / Events

### Text

### Misconceptions

### Interrelationships

### Scale Appropriateness
3. Student Essay and Scoring Keys

Training Examples for Trainer (annotated version)
Training Examples for Trainee
Pre-Scored Essays
  Training essays
  Criterion essays
  Check essays
Essay Scoring Keys
Scored Training Examples for Trainer*

Key to Scale Abbreviations

GICQ: General Impression - Content Quality
PN: Principle/Concept - Number
PK: Prior Knowledge: Facts and Events
TEXT: Text details
MIS: Misconceptions

*Content is underlined and scores are given in parentheses.
The Civil War has been one of the most devastating to the United States. Not just in physical destruction and death, but in the serious ramifications it had to the country's concept of freedom. Some great figures of the period, such as Abraham Lincoln, saw the coming of this, and realized that drastic measures must be taken to keep the nation together regardless of cost. Others, such as Stephen Douglas, believed that the coming war could have been averted by easing tensions between North and South. In his speech Douglas criticized Lincoln's plan to keep the U.S. together by unifying state's dogma; essentially, in the case of slavery. The North, free from slavery, wished to abolish it from the South, while the South wished to keep this institution.

In his speech, Douglas pushed his concept of "popular sovereignty" which was to let each state choose whether slavery was legal in that state. Essentially a lost Federalist, Douglas seemed to have forgotten that the national government was a federalacy not a confederacy as in earlier U.S. history. Because the states of the Confederacy couldn't agree with each other, a new government was drawn up, stated plainly in the Constitution. This was to give the central government power over the states to keep the states essentially united. Although popular sovereignty may have worked, Douglas's problem is that he says in the Constitution each state is sovereign.

Douglas's argument for popular sovereignty is the absence of slavery in such Northern states as New York, Rhode Island, etc., which were slave states originally. Therefore, by letting states be, eventually slavery will die out. Here he gives a bit of information that says, "if it worked here, why won't it work there?" But the problem is that slavery in the South wasn't just a thing on the side; it was a way of life. The south was too deeply rooted in slavery to simply let it die out.
In the year 1776, the colonies of America declared independence from the rule of Great Britain. Several events and debates took place that helped the people of the colonies understand the independence from Great Britain would be the right thing to do.

One such argument against British rule was by Thomas Paine. Mr. Paine stated many opinions about on independence from British rule and what might happen if the American colonies were to stay under Great Britain's government. Mr. Paine states that as a country ruled by Britain, the colonies would be involved in all wars and quarrels between Britain and other countries. These other countries could become allies and friends but instead the colonies would have to fight against the countries. Mr. Paine also wrote that the King of England would still have full say over how the colonies are run. The King of England takes care of England's problems first while the colonies suffer longer and even if help for problems, like war or poverty, arrive, this help would be what ever is left over from Parliament's budget.

There were also several events that helped persuade the colonists to a Declaration of Independence. King George has put several tariffs and laws on the colonies. One such tax is the tea act in which tea has been taxed. The Stamp Act, a tax put on mailing stamps has also outraged the colonists. The worst action is the British soldiers are allowed to sleep and eat anywhere they please and the colonists can not complain. In fact, a brawl broke out in Boston in which several colonists were killed by British soldiers. This is known in the cities as the Boston Massacre.

The debates and events of the past years have helped the colonists of America decide that independence from Great Britain would be better than the continuing British rule over the colonies.
Content Assessment
Training Example 3
Scale: GICQ (1)

We who are of the government of people are writing this information for which you can decide to remain loyal to Great Britain or support the revolution. First, you must know that Britain is the parent country. Europe, but not England is the parent country of America. If any submission or dependence of Great Britain they tend to involve the European Wars, tend to seek friendship, & neither complain, or get angry. Europe is our trade in market. The form of government of Great Britain with serious minds can draw no true pleasure by looking, forward, with pain and positive conviction, which may end sooner or later is what he calls “The Present Constitution” is merely temporary. The king still powers over the government and will have a negative over the whole legislation. American is a secondary object in this system of British politics-England. Independence is the most powerful of all arguments. Independence kept us preserve of the civil wars. You should choice to be interested of Americans to be separated from Britain & neither be reconciliation or independence.

I have learned a lot more that I though I would, but I still feel as though History going to be kinder for me because I don’t enjoy History, but I will give it a shot.
Content Assessment
Training Example 4
Scale: GICQ (2)

The people wanted independence from Great Britain because they wanted to be free and independent. The Americans tried to gain their independence peacefully but it didn't work, they had to war. Their petitions were slighted, their remonstrances produced violence and insult. Some felt, for example, Charles Inglis, if we did fight for our independence, so many lives would be lost, the protection of our trade, all the property throughout the continent would be unhinged. He also thought that it would make this country a scence of blood and slaughter and entail wretchedness and misery. But they decided to fight for their independence.
Yet, how did Lincoln propose to solve this problem? Certainly, if Lincoln's assertion was to be taken seriously then Lincoln indeed ran on a platform of war. With this in mind (in terms of war), two moral issues arose in regards to defense of one's liberties. The first moral issue, of course, was the one of slavery. Slavery was an entirely immoral practice, although it had been practiced for thousands of years and is still employed today. Douglas did not address this, and thus his assertions were incomplete. The other moral issue, so often overlooked by blind critics of Southern conservatives and the like, was now ethical was it to further impose the repressive federal hand upon areas traditionally and properly reserved for the authority of local government. Douglas appropriately addressed this subject by stating the importance of the preserving of local power.
Content Assessment
Training Example 6
Scale: PN (3)

The issues that were dividing the North and South were deeply rooted. It was because of sectional differences such as economy, lifestyle, and the state's rights issue that started the division. The South's plantation economy demanded slaves. The north believed slavery was immoral. The South believed that the individual states could overrule the national government. They believed slaves, which were property, were under the control of the states.
Content Assessment
Training Example 7
Scale: PN (3)

The Declaration of Independence was preceded by many colorful events. Americans pleaded with Americans, and propaganda was everywhere. There were 3 major factors that caused this to happen and these were taxation, King George's wrongs and the writing and speeches of many American authors.

The issue of taxation was probably the biggest cause of friction between England and the colonies in America. Americans never paid their taxes to Britain, yet still claimed they were too high. The Stamp Act was looked at by colonists as dreadful and unjust, but what really happened was Britain cut the taxes in half, but now wanted to collect these taxes. This was interpreted as blasphemy by the colonists, hence the Boston Tea Party, where British tea was dumped into Boston Harbor. The thought of "taxation without representation" also held little water. We wanted to be treated like British, but we failed to realize that we were much better off than the average British citizen. Basically, over "stand" on taxation was falsely based, and unjust.

The next major influence in the Declaration of Independence were the authors and their opinions. They successively influenced the majority, which were indifferent. "Common Sense", by Thomas Pain, was one of the first books to speak out against Britain, and started our move toward independence. He also wrote the Crisis Papers, an extremely melancholy look at the possibility of war. Patrick Henry's speech at the Virginia convention was perhaps the most motivational and inspiring speeches of the times. He addressed the question of independence as that of it being slavery. He made many strong cases against Britain, their increase in arms in the colonies, their blindness to our pleas, and their scornful look upon us. He led us to the point where we must fight for independence, and was so powerful in doing so because of his last line, "I know not what course others may take, but as for me, give me liberty or give me death" This famous line inspired all in the fight for freedom.

The final causes were those of King George's questionable actions. He had sent many troops and arms into the colonies, not, as he said, to protect us, but rather, as Henry said, to control us. This was the first declaration of was, and the colonists also interpreted it that way. George also turned up his nose at our attempts to bargain, and this caused the bitter feelings that culminated in Thomas Jefferson's writing of the Declaration of Independence.

There are many ways to look at the break from England, many different perspectives to see it from. Whatever way a person decides, an America will join in the fight for Independence, and a loyalist will be crushed by the unstoppable power that is the United States of America.
Douglas points out in his speech that states voters must be given the right to determine whether or not to be free or slave. But his argument is struck down by the previous example of Kansas - the Lecompton and Topeka constitution crisis. He fails to realize that people are willing to cheat in order to preserve slavery by voting in other states. And worst of all, he shuns the reality that the people are willing to war against each other as in the John Brown raids to preserve their beliefs. Although the principle of popular sovereignty is fundamentally democratic and allows the voter freedom of choice, the doctrine steers Douglas into the middle ground, no man's land, if you will, where he is subject to attacks by both his party and the pro-slavery Democrats. Popular sovereignty marks a big blow in Douglas' attempt to win the presidency in 1860 against Lincoln.
Because of the policy of "popular sovereignty", the almost newly-formed Kansas became a "bleeding Kansas," people from the north tried to bring people into Kansas to produce a majority, thus making Kansas a free state. This also occurred with the southern people, trying to make Kansas a slave state. As the result of it, riots and small battles took place between the northerners and the southerners.
Content Assessment
Training Example 10
Scale: PK (1)

Black people are just like white people and it is unfair to treat them like property.

Lincoln was from a poor family and was born in Kentucky.

Black were brought to the U.S. from Africa and made into slaves. Often they were treated very badly by their owners.
Content Assessment
Training Example 11
Scale: TEXT (2)

The year is 1858. Politics are becoming more open to the public. Lincoln and Douglas, two men campaigning for Senator of Illinois have just held a debate. They see this country main problems as the possible division of the Northern and Southern states, and how to deal with slavery. Mr. Lincoln said about the division of the union, "A house divided against itself cannot stand." He meant that America cannot survive if it is divided. It needs both the Northern and Southern states working together to keep it prosperous.

Slavery has been a part of our lives for so long that most of us have known nothing else. Some people, they call themselves abolitionists, are tired of slavery. They are trying to change the laws that have made people slaves for so long. The problem is, most southern states are very happy being slave states. They are very heavily into agriculture, and they need people to work in fields. Who can say what's right and wrong? Does congress have the right to decide what every state must do? The country is in a state of turmoil and there don't seem to be too many solutions available. Some people think a war is the best answer, and other want to be passive. No matter what happens, not everyone will be happy with the results.
Mr. Lincoln tells us a house divided against itself cannot stand. He also said I believe this government cannot endure permanently, half slave and half free. We are told that in the future it cannot endure unless they shall become all free or all slave. Mr. Lincoln wishes to go to the Senate of the U.S. in order to carry out that line of public policy.

Mr. Lincoln invites, by his proposition a war of sections, a war between Illinois and Kentucky, a war between free states and slave states. The Difference between Mr. Lincoln and Stephan Douglas upon this point is, that he goes for a combination of the Northern States, or the organizations of the Sectional political party in Free States, to make war until they shall all be subdued and made to conform to such rules as the North shall dictate to them. We agree that, by the constitution we have no right to interfere with it because it is in the constitution, and we are by both duty and inclination to stick by the constitution in all its letter and spirit, from beginning to end.
Through the American Revolution American broke away from England. They did this because the majority felt it was right. American tried to better itself many times.

First, the colonists sent many petitions to England. This didn't affect the British very much. They kept making it difficult for America to prosper.

The Americans also made speeches urging others to rebel. The British did not like this, so they arrested some people and made laws so that others could not keep promoting a rebellion.

The colonists did not like the treatment they were getting. England was taxing them, there were laws that stopped the colonists freedom, and the colonists felt they were not being treated equally. All of this emotion helped to ignite the Revolutionary War.

Throughout the war there were many battles which took place. With the superior army and navy most felt that England would easily win the war, but they had a few obstacles to overcome first. There was the Atlantic Ocean that divided England and America, there was the unfamiliar territory which England was fighting on, and with the ocean news and other things traveled slowly. This made it difficult for the British, and America won its freedom.
Lincoln pretty much tells you in his speech that we should be a free state, free from slavery. Everyone has rights, freedom of speech, why not freedom from slavery? Peace should be spread north and south. Lincoln wanted people from the free states to go south and demand freedom. According to the Kansas-Nebraska (act) bill, there is one possible way to get rid of slavery and that is by leaving a state.

Douglas said the same thing pretty much like Lincoln did about how a house divided against itself cannot stand. He discusses about the subject on slavery like Lincoln did.

Personally, I hate writing assignments that's why I don't do so hot in English. I know that we learn from the past (History) but it's not all that interesting, that's probably why I don't know too much about because I'm more content with the present & future.
Content Assessment
Training Example 15
Scale: MIS (5)

The American Revolution's colonies have just declared their independence from Great Britian. They have fought for numbers of years. One of the main issues they were fighting was about slavery. Slaves wanted to be free. It was a tragedy, brothers fought against each other. It seemed like there was no hope left. A lot of people would die, they would leave the loved ones all alone. The petitions, have been slighted, our remonstrances have produced additional violence and insult, our supplications have been disregarded, and we have been spurned, with concept, from the foot of throne! Patrick Henry was a caring guy, he wanted Great Britian to be friends with us. He thought that there was no reason at all to fight. They should have their own rights like us. He didn't want to see people hurt in any way. All he wanted was no war. Either freedom or death "he said". He would rather live in a free, or die!!! That is a very serious thought. Would you like to live suffering from not being able to do what you wanted? You only live once in your life, try to make the best out of it. But like Patrick Henry, he would rather die, than live an unhappy life. The slaves worked day and night. They worked in fields, and if they weren't where they were supposed to be at the time they were supposed to be there, they would get found and beaten. The is not a way to live. After all the battles in between only the lucky ones did live. Finally, they were free.
One of Douglas's major arguments was for state's rights, he believed, and had evidence to back him up in the Constitution, that states had the absolutely guaranteed right to govern their own local affairs. And this meant that no state had to conform to its fellow states.

He also felt that Northern pride was a tritio. The abolishonists of the Free states never truly wages "political" war against the south, rather they stayed in the North and wages war from there.

He believed that the Northern Republic prejudice was so passionate that it kept Northern scared of making war against southern governments.

Lastly, he felt that if Abolishonists wouldn't have made such a big deal about the slavery issue, gradual emncipation would have occurred in the South as it did in the North. But the hostile reaction of the movements triggered a counter reaction in the South, causing the sectional, racial barrier.
Training Examples for Trainees

Key to Scale Abbreviations:

GICQ: General Impression - Content Quality
PN: Principle/Concept - Number
PK: Prior Knowledge: Facts and Events
TEXT: Text Details
MIS: Misconceptions
Content Assessment
Training Example 1
Scale: GICQ

The Civil War has been one of the most devastating to the United States. Not just in physical destruction and death, but in the serious ramifications it had to the country's concept of freedom. Some great figures of the period, such as Abraham Lincoln, saw the coming of this, and realized that drastic measures must be taken to keep the nation together regardless of cost. Others, such as Stephen Douglas, believed that the coming war could have been averted by easing tensions between North and South. In his speech Douglas criticized Lincoln's plan to keep the U.S. together by unifying state's dogma; essentially, in the case of slavery. The North, free from slavery, wished to abolish it from the South, while the South wished to keep this institution.

In his speech, Douglas pushed his concept of "popular sovereignty" which was to let each state choose whether slavery was legal in that state. Essentially a lost Federalist, Douglas seemed to have forgotten that the national government was a federation not a confederacy as in earlier U.S. history. Because the states of the Confederacy couldn't agree with each other, a new government was drawn up, stated plainly in the Constitution. This was to give the central government power over the states to keep the states essentially united. Although popular sovereignty may have worked, Douglas's problem is that he says in the Constitution each state is sovereign.

Douglas's argument for popular sovereignty is the absence of slavery in such Northern states as New York, Rhode Island, etc., which were slave states originally. Therefore, by letting states be, eventually slavery will die out. Here he gives a bit of information that says, "if it worked here, why won't it work there?" But the problem is that slavery in the South wasn't just a thing on the side; it was a way of life. The south was too deeply rooted in slavery to simply let it die out.
In the year 1776, the colonies of America declared independence from the rule of Great Britain. Several events and debates took place that helped the people of the colonies understand the independence from Great Britain would be the right thing to do.

One such argument against British rule was by Thomas Paine. Mr. Paine stated many opinions about on independence from British rule and what might happen if the American colonies were to stay under Great Britain's government. Mr. Paine states that as a country ruled by Britain, the colonies would be involved in all wars and quarrels between Britain and other countries. These other countries could become allies and friends but instead the colonies would have to fight against the countries. Mr. Paine also wrote that the King of England would still have full say over how the colonies are run. The King of England takes care of England's problems first while the colonies suffer longer and even if help for problems, like war or poverty, arrive, this help would be what ever is left over from Parliament's budget.

There were also several events that helped persuade the colonists to a Declaration of Independence. King George has put several tariffs and laws on the colonies. One such tax is the tea act in which tea has been taxed. The Stamp Act, a tax put on mailing stamps has also outraged the colonists. The worst action is the British soldiers are allowed to sleep and eat anywhere they please and the colonists can not complain. In fact, a brawl broke out in Boston in which several colonists were killed by British soldiers. This is known in the cities as the Boston Massacre.

The debates and events of the past years have helped the colonists of America decide that independence from Great Britain would be better than the continuing British rule over the colonies
Content Assessment
Training Example 3
Scale: GICQ

We who are of the government of people are writing this information for which you can decide to remain loyal to Great Britian or support the revolution. First, you must knew that Britian is the parent country. Europe, but not England is the parent country of America. If any submission or dependence of Great Britian they tend to involve the European Wars, tend to seek friendship, & neither complain, or get angry. Europe is our trade in market. The form of government of Great Britian with serious minds can draw no true pleasure by looking, forward, with pain and positive conviction, which may end sooner or later is what he calls "The Present Constitution" is merely temporary. The king still powers over the government and will have a negative over the whole legislation. American is a secondary object in this system of British politics-England. Independence is the most powerful of all arguments. Independence kept us preserve of the civil wars. You should choice to be interested of Americans to be separted from Britian & neither be reconcilation or independence.

I have learned a lot more that I though I would, but I still feel as though History going to be kinder for me because I don’t enjoy History, but I will give it a shot.
The people wanted independence from Great Britain because they wanted to be free and independent. The Americans tried to gain their independence peacefully but it didn't work, they had to war. Their petitions were slighted, their remonstrances produced violence and insult. Some fell, for example, Charles Inglis, if we did fight for our independence, so many lives would be lost, the protection of our trade, all the property throughout the continent would be unhinged. He also thought that it would make this country a scence of blood and slaughter and entail wretchedness and misery. But they decided to fight for their independence.
Yet, how did Lincoln propose to solve this problem? Certainly, if Lincoln's assertion was to be taken seriously then Lincoln indeed ran on a platform of war. With this in mind (in terms of war), two moral issues arose in regards to defense of one's liberties. The first moral issue, of course, was the one of slavery. Slavery was an entirely immoral practice, although it had been practiced for thousands of years and is still employed today. Douglas did not address this, and thus his assertions were incomplete. The other moral issue, so often overlooked by blind critics of Southern conservatives and the like, was now ethical was it to further impose the repressive federal hand upon areas traditionally and properly reserved for the authority of local government. Douglas appropriately addressed this subject by stating the importance of the preserving of local power.
Content Assessment
Training Example 6
Scale: PN

The issues that were dividing the North and South were deeply rooted. It was because of sectional differences such as economy, lifestyle, and the state's rights issue that started the division. The South's plantation economy demanded slaves. The north believed slavery was immoral. The South believed that the individual states could overrule the national government. They believed slaves, which were property, were under the control of the states.
Content Assessment
Training Example 7
Scale: PN

The Declaration of Independence was preceded by many colorful events. Americans pleaded with Americans, and propaganda was everywhere. There were 3 major factors that caused this to happen and these were taxation, King Georges wrongs and the writing and speeches of many American authors.

The issue of taxation was probably the biggest cause of friction between England and the colonies in America. Americans never paid their taxes to Britain, yet still claimed they were too high. The Stamp Act was looked at by colonists as dreadful and unjust, but what really happened was Britain cut the taxes in half, but now wanted to collect these taxes. This was interpreted as blasphemy by the colonists, hence the Boston Tea Party, where British tea was dumped into Boston Harbor. The thought of “taxation without representation” also held little water. We wanted to be treated like British, but we failed to realize that we were much better off than the average British citizen. Basically, over “stand” on taxation was falsely based, and unjust

The next major influence in the Declaration of Independence were the authors and their opinions. They successfully influenced the majority, which were indifferent. "Common Sense", by Thomas Pain. was one of the first books to speak out against Britain, and started our move toward independence. He also wrote the Crisis Papers, an extremely melancholy look at the possibility of war. Patrick Henry’s speech at the Virginia convention was perhaps the most motivational and inspiring speeches of the times. He addressed the question of independence as that of it being slavery. He made many strong cases against Britain, their increase in arms in the colonies, their blindness to our pleas, and their scornful look upon us. He led us to the point where we must fight for independence, and was so powerful in doing so because of his last line. "I know not what course others may take, but as for me, give me liberty or give me death” This famous line inspired all in the fight for freedom.

The final causes were those of King Georges questionable actions. He had sent many troops and arms into the colonies, not, as he said, to protect us, but rather, as Henry said, to control us. This was the first declaration of was, and the colonists also interpreted it that way. George also turned up his nose at our attempts to bargain, and this caused the bitter feelings that culminated in Thomas Jefferson’s writing of the Declaration of Independence.

There are many ways to look at the break from England, many different perspectives to see it from. Whatever way a person decides, an America will join in the fight for Independence, and a loyalist will be crushed by the unstoppable power that is the United States of America.
Douglas points out in his speech that states voters must be given the right to determine whether or not to be free or slave. But his argument is struck down by the previous example of Kansas - the Lecompton and Topeka constitution crisis. He fails to realize that people are willing to cheat in order to preserve slavery by voting in other states. And worst of all, he shuns the reality that the people are willing to war against each other as in the John Brown raids to preserve their beliefs. Although the principle of popular sovereignty is fundamentally democratic and allows the voter freedom of choice, the doctrine steers Douglas into the middle ground, no man's land, if you will, where he is subject to attacks by both his party and the pro-slavery Democrats. Popular sovereignty marks a big blow in Douglas' attempt to win the presidency in 1860 against Lincoln.
Content Assessment
Training Example 9
Scale: PK

Because of the policy of "popular sovereignty", the almost newly-formed Kansas became a "bleeding Kansas," people from the north tried to bring people into Kansas to produce a majority, thus making Kansas a free state. This also occurred with the southern people, trying to make Kansas a slave state. As the result of it, riots and small battles took place between the northerners and the southerners.
Black people are just like white people and it is unfair to treat them like property.

Lincoln was from a poor family and was born in Kentucky.

Black were brought to the U.S. from Africa and made into slaves. Often they were treated very badly by their owners.
The year is 1858. Politics are becoming more open to the public. Lincoln and Douglas, two men campaigning for Senator of Illinois have just held a debate. They see this country main problems as the possible division of the Northern and Southern states, and how to deal with slavery. Mr. Lincoln said about the division of the union, “A house divided against itself cannot stand.” He meant that America cannot survive if it is divided. It needs both the Northern and Southern states working together to deep it prosperous.

Slavery has been a part of our lives for so long that most of us have known nothing else. Some people, they call themselves abolitionists, are tired of slavery. They are trying to change the laws that have made people slaves for so long. The problem is, most southern states are very happy being slave states. They are very heavily into agriculture, and they need people to work in fields. Who can say what’s right and wrong? Does congress have the right to decide what every state must do? The country is in a state of turmoil and there don’t seem to be too many solutions available. Some people think a war is the best answer, and other want to be passive. No matter what happens, not everyone will be happy with the results.
Content Assessment
Training Example 12
Scale: TEXT

Mr. Lincoln tells us a house divided against itself cannot stand. He also said I believe this government cannot endure permanently, half slave and half free. We are told that in the future it cannot endure unless they shall become all free or all slave. Mr. Lincoln wishes to go to the Senate of the U.S. in order to carry out that line of public policy.

Mr. Lincoln invites, by his proposition a war of sections, a war between Illinois and Kentucky, a war between free states and slave states. The difference between Mr. Lincoln and Stephan Douglas upon this point is, that he goes for a combination of the Northern States, or the organizations of the sectional political party in Free States, to make war until they shall all be subdued and made to conform to such rules as the North shall dictate to them. We agree that, by the constitution we have no right to interfere with it because it is in the constitution, and we are by both duty and inclination to stick by the constitution in all its letter and spirit, from beginning to end.
Through the American Revolution American broke away from England. They did this because the majority felt it was right. American tried to better itself many times.

First, the colonists sent many petitions to England. This didn't affect the British very much. They kept making it difficult for America to prosper.

The Americans also made speeches urging others to rebell. The British did not like this, so they arrested some people and made laws so that others could not keep promoting a rebellion.

The colonists did not like the treatment they were getting. England was taxing them, there were laws that stopped the colonists freedom, and the colonists felt they were not being treated equally. All of this emotion helped to ignite the Revolutionary War.

Throughout the war there were many battles which took place. With the superior army and navy most felt that England would easily win the war, but they had a few obstacles to overcome first. There was the Atlantic Ocean that divided England and American, there was the unfamiliar territory which England was fighting on, and with the ocean news and other things traveled slowly. This made it difficult for the British, and America won its freedom.
Content Assessment
Training Example 14
Scale: MIS

Lincoln pretty much tells you in his speech that we should be a free state, free from slavery. Everyone has rights, freedom of speech, why not freedom from slavery? Peace should be spread north and south. Lincoln wanted people from the free states to go south and demand freedom. According to the Kansas-Nebraska (act) bill, there is one possible way to get rid of slavery and that is by leaving a state.

Douglas said the same thing pretty much like Lincoln did about how a house divided against itself cannot stand. He discusses about the subject on slavery like Lincoln did.

Personally, I hate writing assignments that’s why I don’t do so hot in English. I know that we learn from the past (History) but it’s not all that interesting, that’s probably why I don’t know too much about because I’m more content with the present & future.
Content Assessment
Training Example 15
Scale: MIS

The American Revolution's colonies have just declared their independence from Great Britian. They have fought for numbers of years. One of the main issues they were fighting was about slavery. Slaves wanted to be free. It was a tragedy, brothers fought against each other. It seemed like there was no hope left. A lot of people would die, they would leave the loved ones all alone. The petitions, have been slighted, our remonstrances have produced additional violence and insult, our supplications have been disregarded, and we have been spurned, with concept, from the foot of throne! Patrick Henry was a caring guy, he wanted Great Britian to be friends with us. He thought that there was no reason at all to fight. They should have their own rights like us. He didn't want to see people hurt in any way. All he wanted was no war. Either freedom or death "he said". He would rather live in a free, or die!!! That is a very serious thought. Would you like to live suffering from not being able to do what you wanted? You only live once in your life, try to make the best out of it. But like Patrick Henry, he would rather die, than live an unhappy life. The slaves worked day and night. They worked in fields, and if they weren't where they were supposed to be at the time they were supposed to be there, they would get found and beaten. The is not a way to live. After all the battles in between only the lucky ones did live. Finally, they were free.
Content Assessment
Training Example 16
Scale: MIS

One of Douglas’s major arguments was for state’s rights, he believed, and had evidence to back him up in the Constitution, that states had the absolutely guaranteed right to govern their own local affairs. And this meant that no state had to conform to its fellow states.

He also felt that Northern pride was a trait. The abolitionists of the Free states never truly waged “political” war against the south, rather they stayed in the North and waged war from there.

He believed that the Northern Republic prejudice was so passionate that it kept Northern scared of making war against southern governments.

Lastly, he felt that if Abolitionists wouldn’t have made such a big deal about the slavery issue, gradual emancipation would have occurred in the South as it did in the North. But the hostile reaction of the movements triggered a counter reaction in the South, causing the sectional, racial barrier.
Dear cousin, the debate was between Lincoln and Douglas; they are running for Senate.

Dear cousin, the debate was between Lincoln and Douglas; they are running for the Senate for the state of Illinois.

Douglas claims that Lincoln is trying to start a Civil War with the South. He says we've been divided into Free States and Slave States from 1789. Now Lincoln says we cannot anckle in the
future unless the Republic is all free or all slave. Douglas claims
Lincoln invites the South to make
warfare upon the North and
establish slavery. Therefore
Lincoln invites by his words
a war between North and South,
a war between brothers. "One
section or the other shall be
entirely subdued," said Douglas
trying to show how Lincoln
wanted war.

Lincoln comes up to

debate his side, and says, "I
made prediction only -- it may have been a foolish one, perhaps. "Referring to his recent speech at Springfield were he said, "A house divided against itself cannot stand. " Lincoln believes the only reason slavery has lasted so long is, because people thought that slavery was in the course of ultimate extinction."

"The Constitution decree that slavery should not go into the new Territory," said Lincoln. "If, claim he is not true.
to start a war between the 
North and South. He claims 
Douglas sees slavery as a 
very little matter, and he 
hates slavery.

Personally I think 
Lincoln was right slavery 
is wrong, but what can be 
done. I don't want a war 
over the issue, because some 
day it will go away. Lincoln 
is trying to start a war 
between us and them. When 
I don't know who us and
them are. Sure its Northern States against Southern, but which am I? How can we fight against our families and friends on the other side? I hope Douglas is wrong and there is no war. On the other hand I hope Lincoln is right, and slavery is ended.
Essay

In 1858 Abraham Lincoln and Stephen Douglas campaigned for the office of Senator from the state of Illinois, they held seven joint debates throughout the state. Mr. Lincoln tells in Springfield before the convention which gave him his unanimous nomination, that "a house divide against itself cannot stand." Which means that the country in separate parts against each other will not last. And, "I believe this government cannot endure permanently half free and half slave." Which means that the country can't live with half free, half slaves. "It will become one or the other." Although this country has existed from 1789 to this day divided into free states and slave states, yet we are told that in the future it cannot endure unless they shall become all free or all slave.
He wishes to go to the Senate of the United States in order to carry out that line of public policy which will compel all the States in the South to become free.
1. American history teaches us many things in which we can learn from what has happened. It has changed a lot of things. There been the movement of rights, slavery, freedom and the list can go on and on. The debates between Abraham Lincoln and Stephan Douglas had some differences. Abraham Lincoln feels that the combination of the Northern States, on the organization of a sectional political party in the Free State to make war until they shall all be subdued, and made to conform to such rules as the North shall dictate them. Their war, a making war, is not to assert
into those States where slavery exists, and there interfere, and render themselves responsible. Douglas is opposed to a sectional agitation to control the institutions of other States. He is opposed to organized a sectional party, and which appeals to northern pride, and northern passion and prejudice against Southern Institutions. Slavery is the issue here in which both groups have different views: at it, there is only one way in which it could be abolished and that is by leaving a state according to the Nebraska-Bill, meaning to settle. Have to form and populate it.
Institutions in its own way.

Abraham stated that Douglas made accusations which weren't true. He said that Abraham was in favor of all these things like making war between North upon the South for the extinction of slavery; inviting the South to a war upon the North for the purpose of nationalismizing slavery. And what Abraham was doing was making only predictions. The constitutions is what is nothing the work together it states rules in which people have to obey and if they don't they not punishment. What Abraham said is
that slavery was a big issue and Douglas thinks it's a mere little thing. They look upon it as a vast moral evil.

With reading both of the debates there was much to be said about slavery. I feel that slavery was wrong; nobody mean no one should be treated badly unless they do bad. And punishment should be upon that person to a certain extent. Slavery itself was very unnecessary, and it caused hardships between people. I'm so glad that
Slavery is over with now and everyone is treated with equality and respect.
The debate was quite heated, and poignant speeches were made. Lincoln and Douglas verbally fought each other on the way through Illinois. At the debate I went to, Stephen Douglas seemed more interested in picking apart quotes and phrases by Lincoln than further explaining his platform. He did not mention the horrors of slavery—the abuse, the maltreatment, the humiliation. Douglas is a product of a wealthy family who never let him see what the bad side of town looked like—and he prefers to keep it that way. His opinion of upper and lower class, and his contempt for black race shows his ineptitude as a politician. His goal is to reach all people, not just the wealthy merchant class, who appeal to material desires. In Douglas' speech, he makes the point that only a ratification of the Constitution would change slave states to free ones. My question to him was, what would be so terrible about changing the Constitution especially to correct a situation that is ethically and morally wrong?
Lincoln, however took Douglas’ snappish, desperate remarks in stride, and quietly commented on his own speech; in effect, clarifying what he’d said earlier and making Douglas look like a nit-picking fool. Douglas also assumed too much, as Lincoln says—

"...In this paragraph, Judge Douglas thinks he discovers great political heresy. I want your attention particularly to what he has inferred from it. He says I am in favor of making the whole States of this Union uniform in all their internal regulation; that in all their domestic concerns I am in favor of making them entirely uniform. He says I am in favor of making war... for the extinction of slavery; that I am also in favor of inviting (as he expresses it) the South to war... for the purpose of nationalizing slavery. Now... if you will carefully read that passage... I did not say that I was in favor of anything in it."

Judge Douglas jumped to conclusions, stating that the only way to abolish slavery was to close state legislature and have Congress make all domestic laws and regulations. Lincoln’s reasons for abolishing slavery are pure and simple; he believed it was morally wrong.
One of the main facts, and one that I think is very important is the issue of slavery. 

Steven Douglas mocked Abraham Lincoln on a very brilliant speech. He said, "A house divided against itself cannot stand." This brings up a major point in saying that when the southern states are allowing slavery, the northern states are beginning to feel hatred against the South. Steven Douglas also mishandled or else misinterpreted Lincoln's speech telling all who would listen
that Lincoln was starting a war between the states.

Now, don't get me wrong, Lincoln defended himself. He did so in a very unique way (As speaking in terms of the government), in that he did not once mock or deride Steven Douglas to any low standards. He simply reexplain his first speech so that no person could misunderstand him again.

This, I thought, made him a good spot.

For the most part, I agreed with Lincoln in the fact that our nation...
needs either slavery for all states or slavery for none. A war would not settle this dispute, because, I think if a war would eventually create the constitutional bond between the states to loosen and possibly fall apart creating two nations inside of the United States.

As the constitution states, "We the People..." as in a whole need to settle this without violence and without hate. Lincoln is willing to do this.
And to conclude, I'd like to say that in the debate of Douglas-Lincoln, I truly believe in Lincoln's speech as the best. This is the one that believes in the Constitution from beginning to end....
The problem facing the nation now is whether America can exist half slave and half free. Douglas seems to believe this is possible through the fact that it has been that way from 1789 until now. Lincoln feels the government has held together that long until the Nebraska bill was introduced.

America was beginning to crumble as a union. The North and the South had many opposing viewpoints on the issue of slavery and it began a large dispute between the two. Lincoln felt America could not exist half slave half free.
yet Douglas argued that what had been going on since 1789.
Lincoln quotes in his speech "I believe this Government cannot endure permanently half slave and half free.
issue was not how long the government had stood, but what is was only being brought down because of the Fugitive Slave Bill. He predicted that "slavery was only in cause of ultimate extinction. He hated slavery and felt that the Constitution backed up his belief in "ultimate extinction" of slavery by pointing out the fact that slavery would be prohibited in the new territory and that the African Slave Trade would be cut off by Congress.
Douglas, however, felt that Lincoln was beginning to propose a war. He believed that America could exist half slave and half free since they had been in so long. He thought that there would be no peace between the North and the South if sections were forced to abolish slavery. He believed the Constitution preserved the liberties, rights, and sovereignty of each state and said that it was not possible to "establish uniformity in the local and domestic institutions of all the states." He also believed that the only way to abolish slavery would be to change the Constitution which would take away the rights of the states.
There are many problems in the United States, but the two greatest seem to be the issues of slavery and sectionalism. These two problems go hand in hand, for slavery actually intensifies sectionalism. Sectionalism has existed since the beginning of our nation, but it has grown progressively worse since the issue of slavery came to a head.

Slavery has literally divided the nation in half. Most, although not all, Northerners are opposed to it, and most Southerners are in favor of it. These divisions are based on the very different economies that exist in the North and South. The North is basically industrial, with a denser population and more factories. The South, on the other hand, produced raw materials that are usually grown on large plantations. The whole Southern economy rests on the slavery issue, and it faces a possible collapse if the slaves are freed, because there will be no one to work the fields.

One possible solution to the problem is gradual emancipation. This would perhaps save the South from the economic collapse, while satisfying the abolitionists who want slavery done away with. Another solution would be to simply divide the United States into two separate countries, one slave and one free. This, however, would produce two countries that would be much weaker both economically and defensively, than one united country. Still another way of dealing with the issue is to have a war. If the free side wins, then the United States will be free, if the slave side wins, then slavery will remain legal. If the latter happens, then the circle of problems and sectionalism might begin all over again.

The slavery issue has intensified the problem of sectionalism. Many leaders in the South want a weaker federal government, and stronger state governments. They want each state or territory to have the power to decide whether it will be slave or free.
If the state governments become too powerful, though, then each state will be competing against the others and we will have a nation divided into many different countries. A strong federal government would mean a stronger nation as a whole. Although sectionalists are concerned with the problems of one nation, these problems are looked at on the local level instead of the national level. This is almost like selfishness, because it weakens the country to have dissension within it.

The federal government cannot be allowed to become too strong, however, because soon the people would have no say at all in the government. There needs to be a balance of power between the state and federal governments.

The United States is now facing a civil war if certain serious problems are not solved. The issue of slavery must be decided, as well as the problem of sectional rivalry. If a solution to these problem is not found, then the United States of America might self-destruct from within.
Mr. Lincoln said, and I quote: "A house divided against itself cannot stand. I believe government cannot endure permanently, half slave and half free." "I do not expect the Union to be dissolved, I do not expect house to fall; but I do expect it will cease to be divided. It will become all one thing or all the other."

That is the purpose and principle upon which he sets out on this campaign. This has all been quoted and requested by him. Most Lincoln.
have also been quoted by Stephen A. Douglas.

The war between the north and south is somewhat for the freedom of slaves. The Constitution and what it reads about freedom and slaves is always being discussed in Lincoln and Douglas's speeches.

Slavery is probably the main fact here besides the fact of the war. I am to understand that Lincoln is not in favor of slavery.
So far as the north is not, Mr. Douglass believes that there is but one possible way in which slavery can be abolished, and that is by freeing a state, according to the principle of the Kansas-Nebraska bill, perfectly free to form and regulate its institutions in its own way. It is under this principle in which he wishes to proceed. He does not believe in slavery. Mr. Lincoln believes that we cannot eliminate slavery.
in the states one by one
best by deciding all together.
also stating the Constitution in the
last words of the debate Lincoln
says: 'we have no right to interfere
with it (the Constitution), because it is
in the Constitution and we are by
folk duty and inclination to stick by
that Constitution in all its letters and
spirit, from beginning to end.'
I. 1858: Lincoln vs. Douglas, 2 problems
   A. slavery itself
   B. nationalism vs. sectionalism

II. slavery itself
   A. Since 1785
   B. South: dependent
   C. North: no use
   D.
At the time when Lincoln and Douglas were debating for the office of Senator in 1858, the United States faced several problems that were very often closely related to each other. One of the main problems was slavery, and another major issue was the struggle between sectionalism and nationalism.

Slavery had been in the United States for a long time. The Spanish conquistadors enslaved Indians; the British imported blacks from Africa later. While the South of the U.S. had a demand for black slaves (hot weather huge plantations, hard work), the North had no use for them (cold weather is unfavorable to blacks). Therefore, the Union was soon divided into free states and slave states. For years, there had been a balance between the powers of free and slave states; however, as the frontier was pushed westward, it soon became obvious that there would be problems about "alignment" of newly founded states. Debates, discussions, regulations were made to regulate which of the new states were to be free and which were to be slave states. An example was the Missouri Compromise, which drew a line through the U.S. from East to West; south of this line, slavery was tolerated, north of it it was forbidden. Another example is the Kansas-Nebraska Act, which let states decide on their own (by popular vote). However, sooner or later the slave states would have been outnumbered and therefore lost power in Congress. This led eventually to the Civil War.

Another problem closely related to slavery was the struggle of nationalism against sectionalism. Nationalists were in favor of a strong federal government with much power over the individual states.
Sectionalists wanted the opposite: a loose confederation rather than a strict Union, a weak federal government and strong individual states (states rights). As Northerners (who controlled the majority of the Congress) tried to restrict slavery and started laws against it, Southerners questioned that the government had the right to do it. They said it should be left to each individual state to choose between slavery or abolition; this was not supposed to be the decision of the central government. This thinking also led to secession of the Southern states, which then founded a confederation.

These two problems - slavery and sectionalism - finally led the U.S. into a civil war. No side was, in the end, willing to compromise (especially on the issue of slavery). It would probably have been best to let the issue simply settle down and let each state decide on its own. This could have been able to prevent the war.
The main problem facing America at this time was the issue of slavery and the seceding of the southern states. The issue of slavery was brought up many times during the Lincoln-Douglas debates. Lincoln was antislavery, but didn't state that very loudly. Douglas was also antislavery, but liked the idea of popular sovereignty for territories wanting to join the union. Lincoln wanted either slavery or no slavery, no inbetween.

Douglas thought that the states should do whatever they want, but slavery will eventually be ended by the souther...
state. He thought this would take time but it was worth it to save the United States from a civil war.

The separation of the union was looked upon as a threat to the union so he sent soldiers to try to win Richmond from the South but he was surprised when his northern troops came running back to Washington with their tails between their legs. The Rebel force decided not to chase the retreating army back North, which proved to be a good idea. Stephen Douglas said the...
"he was opposed to the forming of a sectional party, which appeals to northern pride, and northern passion and prejudice, against southern institutions, thus stirring up ill-feeling and hot blood between brethren of the same Republic." This made him impartial to the separation of north and south.

These debates were very informative about Douglas's inability to comprehend what was happening in the south.
Lincoln vs. Douglas were arguing over the issue of slavery, whether it should be abolished right away or let it become extinct on its own. Lincoln said, "A house divided against itself cannot endure permanently, half slave and half free." Douglas believed that if the union survived another divided against itself, it could endure until slave became extinct on its own. Douglas said that Lincoln was pushing toward war, but in reality Lincoln was against war. It was necessary to preserve the union then.
Lincoln's main concern was to preserve the union. Douglas just kept arguing that Lincoln was leaning toward war. Lincoln was open to anything that could restore peace between the North and the South, "if such a Doctrine should prevail in either section of the Union!"

Once the Nebraska bill was accepted, the citizens minds were at rest believing that slavery was in the state of ultimate extinction.

Lincoln was also hurt by the fact that the federal government couldn't
enforce laws in the South as long
as the South did want them to.
Douglas said that laws from one
state are not going to be enforced in
another state. For instance, "The right of
Illinois don't interfere with the cranberry
laws of Indiana, the oyster laws of Virginia,
or the liquor laws of Maine.

Douglas continually stated his
feeling that slavery was in the state of
ultimate extinction. He said that "Under the
Repubic's operation that the union has survived,"
"Under its operation slavery has disappeared
in New Hampshire, Rhode Island, Connecticut,
New York, New Jersey, Pennsylvania, from six of the twelve original slaveholding States.

"The gradual system of emancipation went on quietly, peacefully, and steadily, as long as we in the free States minded our business and left our neighbors alone." Douglas said.

In effect, abolition societies brought up violent crusades against the South, and this brought violent uprisings from the South. Lincoln and Douglas argued this point back and forth, until Douglas won the seat in Senate, but Lincoln had the ultimate victory by becoming President at a later time.
In this period of time there were not so good presidents. Then Abraham Lincoln became the president. He was a good president and helped his country out a lot. He issued the Emancipation Proclamation and he also kept England out of the Civil War. There was hatred also brewing in this time period. The South and the North were either going to have to become either all slave states or all free states. Then the South fired on Fort Sumter. By then firing on Fort Sumter this made a friend of Lincoln write a note to Lincoln to come home.
and supplies. By this Lincoln had a call up to arms. And by Lincoln sending the troops and supplies the Upper South finally succeeded from the Union. The Lower South had already succeeded from the Union because Lincoln was elected president and then Lincoln's April Policy made the Upper South succeed. This broke out as a Civil War. The North vs. the South. At the beginning of the war the South had a lot of good generals, generals like Robert E. Lee and Stonewall Jackson and the North had men like Phil. Burnside and...
McClellan. Robert E. Lee was an outstanding general. He only went on the offensive side three times and won once but on the defensive side he was just about unstoppable.

He was one of the greatest generals in the history of the United States. "Stonewall" Jackson was also another great Southern general. He was the greatest outflanking general in the world. The North had Hooker, Burnside, and McClellan. They weren't smart at all. Even though McClellan had the army of the Potomac he didn't know how to use it. Then the North finally got generals like Grant, Sherman and Meade. Then there...
Also one of the best generals the US had produced. Finally with the combined power of these men they over took the South and Robert C. Lee's army. The Civil War ended at Appomattox Court House. Lincoln was assassinated about 6 weeks after this at Ford's Theatre.

I learned a lot from the reading yesterday. I learned the different views of Lincoln and Douglas about slavery and the issue between the North and South. Lincoln's ideas about this were that a house divided against itself cannot stand and
that this government cannot endure permanently, half slave and half free, that the Union will divide and it will become all one thing or the other. He said that we must preserve the Union and that there is one way that slavery can be abolished and that is by having a state choose their own way. Stephen Douglass looks at slavery as a small thing and really doesn't bother with it and then he says that he thinks that those want to or is in favor for having the North go into the South and -
Abraham Lincoln believed that every state should be free. He didn't think it should be half and half. He wanted to fight for all states being free or all having slavery because a house divided against itself cannot stand. Abraham Lincoln hated slavery so of course, he wanted free states. He wanted to put an end to slavery. I don't know much about the Civil War, but I think this in the way it came about
Because Mr. Lincoln wanted only Free States or only Slave states a war broke out. The Civil War. In the meantime Slaves were being rescued to the North through a route called the Underground Railroad. The Underground Railroad really wasn't underground it was just a series of houses all the way from the South to the North.

After a while of bloodshed slavery was gone. All the
States were in unity and there was no more "North and South"
1) I am a middle aged banker who has lived in Illinois most of my life. I have a wife and two small children.

2) Kansas-Nebraska bill, Cranberry laws of Indiana, oyster laws of Virginia, Liquor laws of Maine

3) The sovereignty of state law, the fundamental principles of the Constitution

4) Slavery, Sectionalism

Essay

"We can talk in the parlor, John." Henry placed an arm upon his cousin's shoulder and began walking out of the living room and into the dim light of the parlor. "Have a seat John, can I get you a drink?"

John carefully sat himself down in the chair nearest the window. "Thank you Henry, brandy will be fine." Lightning flashed outside the window and rain pelted it with an ever increasing intensity. "So, what's all the new commotion in the states about Henry? We don't hear many details from across the ocean you know?"

Henry poured two brandys and handed one to John. He seated himself next to his cousin and felt the heaviness of his legs, the trip to the debate had been long and he was quite tired. "Well to tell the truth I feel a bad time for the Union coming along." He sipped at the brandy and let it warm over his insides. "For instance, I took Emily and the children to the Lincoln Douglas debate today, seems like the topic of the day was slavery."

"Ah, yes. I had heard something about that in England."

"Well Douglas believes that if Lincoln were to abolish slavery the Southern states would never agree to it willingly. Douglas also thought..."
that it would be better to let slavery run its course and die out by itself. Between sips of brandy and pauses for the bellowing sounds of thunder he continued "Lincoln, and a lot of other people on his side, believe that it is time slavery was taken out of the Union. He preached about the moral indignity's and all that."

"Well what does the constitution say?"

"Well Douglas threw the constitution into it by bringing up the point that the federal government is not supposed to interfere with states rights. But Lincoln thought that the Constitution was written under the precept that someday slavery would be done away with." Henry sat his glass down on the table and continued "People are also getting mighty afraid that the issue of slavery is gonna drive a wedge between the North and South. That's just the kinda thing that scares me. Seems as time goes by everyone gets more pride in the direction they live in than the country they live in. No one wants to be an American anymore, everyone wants to be a Northerner, or a Southerner."

John set his glass next to Henry's and stood up to look out the window. "So Henry, what do you think it'll all come down to?"

"Well... to tell ya the truth, I think Lincoln's right. We can't stand as a nation divided within itself. Just because we've done it so far doesn't mean nothin! The way folks are all riled up over this I think slavery should be abolished no matter what the cost. It's about time somebody like Lincoln came around. I just hope he can do it without pulling the country apart at the seems. I feel a bad storm coming up for the Union, cousin.... a bad storm."
## Training Essay Scores

**Topic Key**  
L/D = Lincoln/Douglas  
H/I = Henry/Inglis  
P/I = Payne/Inglis

<table>
<thead>
<tr>
<th>Student ID</th>
<th>Rater</th>
<th>Topic</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>L/D</td>
<td>GICQ</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>L/D</td>
<td>PN</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>L/D</td>
<td>PK</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>L/D</td>
<td>TEXT</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>L/D</td>
<td>MIS</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>L/D</td>
<td></td>
<td>2</td>
<td>3</td>
<td>5</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>7</td>
<td>L/D</td>
<td></td>
<td>5</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
## CRITERION ESSAY SCORES

### TOPIC KEY
- L/D = LINCOLN/DOUGLAS
- H/I = HENRY/INGLIS
- P/I = PAYNE/INGLIS

<table>
<thead>
<tr>
<th>Student ID</th>
<th>Rater</th>
<th>Topic</th>
<th>1 GICQ</th>
<th>2 PN</th>
<th>3 PK</th>
<th>4 TEXT</th>
<th>5 MIS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>L/D</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>L/D</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>L/D</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>4</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Student ID</td>
<td>Rater</td>
<td>Topic</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>------------</td>
<td>-------</td>
<td>-------</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>1</td>
<td>L/D</td>
<td>G1C0</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>L/D</td>
<td>PN</td>
<td>2</td>
<td>1</td>
<td>5</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>L/D</td>
<td>PK</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>L/D</td>
<td>TEHT</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>