

**CRESST Performance Assessment Models:  
Assessing Content Area Explanations**

CSE Report 652

Eva L. Baker, Pamela R. Aschbacher,  
David Niemi, and Edynn Sato  
CRESST/University of California, Los Angeles

April 1992

Center for the Study of Evaluation  
National Center for Research on Evaluation,  
Standards, and Student Testing  
Graduate School of Education & Information Studies  
University of California, Los Angeles  
Los Angeles, CA 90095-1522  
(310) 206-1532

Eva L. Baker, Project Director, CRESST/UCLA

Copyright © 1992 The Regents of the University of California

The work reported herein was supported in part under the Educational Research and Development Centers Program, PR/Award Number R305B960002, as administered by the Institute of Education Sciences, U.S. Department of Education and R117G10027.

The findings and opinions expressed do not reflect the positions or policies of the National Center for Education Research, the Institute of Education Sciences or the U.S. Department of Education.

This report was published as a guidebook in 1992 and has not been revised. However, it represents a significant CRESST research publication.

# Table of Contents

**Introduction** ..... 3

**Chapter 1**  
Overview of CRESST Research ..... 5

**Chapter 2**  
Guidelines for Using CRESST's Model  
for Assessing Explanation ..... 9

**Chapter 3**  
Sample Assessment Materials for Students ..... 16

**Chapter 4**  
Specifications for Developing Assessment Materials ..... 33

**Chapter 5**  
Rater Training, Scoring and Reporting ..... 41

**Chapter 6**  
Sample Training Materials ..... 66

# Introduction

## *Purposes*

This handbook presents a performance-based approach to assessing students' understanding of subject matter content. It is based on years of research conducted by the National Center for Research on Evaluation, Standards, and Student Testing (CRESST), funded by the U.S. Department of Education's Office of Educational Research and Improvement (OERI). The purposes of this handbook are to:

- provide one model of alternative assessment for those who need to develop similar assessments of their own;
- introduce successful examples of CRESST assessment materials; and
- facilitate research on other alternative assessments.

The materials which follow are the result of our five year research effort designed to explore the development of alternative assessments in history. To summarize, the project has attempted to find ways to score the content quality of essays in history. Using the writing of expert historians as the basis of scoring criteria, we have developed techniques for measuring the deep understanding of history and for scoring student work reliably. Our work has been conducted using students from grades 8 through 12 and has been expanded to other content areas as well (economics and science).

## *Large-Scale Assessment*

These assessment tasks are consistent with cognitive learning theory. They include recalling prior knowledge in a content area, reading primary source documents containing new information, and writing an explanation of important issues that integrates new and prior information.

Our assessment judges student understanding on the basis of six scales, including the use of concepts and facts, the avoidance of major misconceptions, and the quality of the argument presented. The scales were developed from studies of expert and novice performance. We have used this assessment approach to research a number of technical issues in performance assessment and have demonstrated the reliability, validity, and generalizability of this technique.

***Reliability,  
validity, and  
generalizability***

We believe that this assessment could be useful for both large-scale accountability and 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 they can be merged into a single measure, with different methods of reporting the data for different purposes.

Inside you will find background information on our CRESST performance-based assessment, examples of assessments for secondary level history and chemistry, and specifications for duplicating our technique with other topics and subject matter areas. We also describe our rater training process, scoring techniques, and methods for reporting results.

Interested users may contact CRESST at (310) 206-1532 for copies of additional materials, assistance using them in an assessment program, help in developing assessments for new topics, or for technical information about the rating scales.

# Chapter 1

## Overview of CRESST Research

### Cognitively Sensitive Assessment

Tests should measure significant learning in a way that supports desired performance. This simple concept should lead us, as educators, to a reversal of our present use of standardized tests which fail to measure deep understanding of student learning. Instead of having tests constrain instruction, assessment procedures should build directly on learning.

Despite the widespread interest in alternative assessments, there has been relatively little research on the design and technical quality of such measures. CRESST began conducting research on history performance measures in 1988. Focusing on both explanation and knowledge representation skills, we have attempted to develop a better method for validly assessing secondary students' deep understanding of content areas such as history.

Many current performance assessments are developed with minimal design constraints because clearly acknowledged technology does not exist for performance task design. Developers seem to focus on a few limits when they create new assessments. One set of constraints concerns logistical issues, such as assessment time and availability of materials. Another emphasis has been on the surface characteristic of the task, that it exhibits motivational or "authentic" attributes of the assessment.

Teachers and other developers want assessments that capture the imagination of students, intrinsically motivate, and if possible, demonstrate relevancy to real-world demands and expectations. Far less attention has been paid to design constraints focused on increasing the technical quality and the economic feasibility of the resulting assessments.

*Assessment procedures should build directly on learning*

*Assessments that capture the imagination of students*

***Control both  
rater and score  
reliability***

***An essay that  
explains the  
positions of the  
authors***

CRESST's research assumes that a desired goal of performance assessment is the generation of "comparable" tasks for estimating student achievement. Our approach has sought to produce comparability by designing it at the outset rather than adjusting findings *post hoc* through scaling and statistical equating. Specifications to control cognitive demands of the task, the structure of the assessments, and the generation and application of scoring rubrics have been thought to produce performance that showed less variability from topic to topic than tasks created with fewer design constraints. In our attempt, we have tried to control both rater and score reliability.

Our history performance tasks, which have evolved over time, require students to engage in a sequence of assessed steps—taking a minimum of one-and-a-half hours per topic. First, students are assessed on their relevant background knowledge of the particular historical period. This measure consists of a 20-item, short-answer test with questions to measure student knowledge of historical principles and specific events pertinent.

Next students are provided with opposing viewpoints in primary source text materials, typically letters or speeches of historical figures. Finally, students are asked, in a highly contextualized set of directions, to write an essay that explains the positions of the authors of the texts, and to draw upon their own background knowledge for explanation. In some studies we have given students optional resources to read, or have asked students to prepare HyperCard or concept map representations of the key knowledge, principles and relationships in the text materials (Baker, Niemi, Novak, & Herl, in press).

CRESST conducted a series of studies to determine how scoring rubrics should be developed, and the best strategy relied on looking at differences between expert and novice performance (Baker, Freeman, & Clayton, 1991). The essay scoring rubric consists of six dimensions, a General Impression of Content Quality scale (focused on the overall quality of the content understanding), and five analytic scales:

***Five analytic scales***

- Prior Knowledge (the facts, information, and events outside the provided texts used to elaborate positions);
- Number of Principles or Concepts (the number and depth of description of principles);
- Argumentation (the quality of the argument, its logic and integration of elements);
- Text (the use of information from the text for elaboration);
- Misconceptions (the number and scope of misunderstandings in interpretation of the text and historical period).

Each of the above dimensions is scored on a 0-5 point scale.

***Sets of history assessments have been developed***

History experts and high school teachers have been involved throughout the study as co-designers, reviewers, and raters of the assessment. So far, six complete sets of history assessments have been developed: two on the Revolutionary period; one on the Civil War; two on 20th century immigration; and one on the Depression Period. These tasks connect to the *California History-Social Science Framework* (1988). Replications in the areas of science (Baker, Niemi, Novak, & Herl, in press) and economics (Baker, 1991) have been conducted to assess the utility of the scoring rubric for explanation tasks in other content areas.

**What CRESST Has Learned**

Over the past several years of research on this project, CRESST has:

***CRESST research***

1. developed a valid scoring scheme for assessing deep understanding of history, generalizable across topics;
2. developed rater training procedures that produce reliable and valid scoring of student tasks in a limited period. The scoring rubric makes strong cognitive demands of the raters;



***What CRESST  
has learned***

***Fairness,  
generalizability,  
cognitive  
complexity,  
content quality,  
reliability,  
cost, and  
efficiency***

3. built a task structure that reduces score variability so that fewer topics can be used to derive reliable scores for individual students. This technique is more efficient than found in most comparable studies. These relationships are all the more startling because of the lack of preparation and motivation among our students;
4. distinguished between assessment purposes and the utility of overall score and subscores;
5. found gender differences in this small sample, favoring females;
6. found supportive data for the validity of our measures in grade point average (GPA) and a scale measuring student effort;
7. systematically addressed validity criteria (Linn, Baker, & Dunbar, 1991) in our research studies: the criteria addressed include fairness, generalizability, cognitive complexity, content quality, reliability, cost and efficiency. We are in the process of conducting studies of transfer and designing research to assess the meaningfulness of tasks to students.

For additional details on the background, development and methodology of this research, please contact the CRESST office.

**References**

- Baker, E.L. (1991). *Foundation for Teaching Economics evaluation report*. Los Angeles, CA: Author.
- Baker, E.L., Freeman, M., & Clayton, S. (1991). Cognitive assessment of history for large-scale testing. In M.C. Wittrock & E.L. Baker (Eds.), *Testing and cognition*. Englewood Cliffs, NJ: Prentice-Hall.
- Baker, E.L., Niemi, D., Novak, J., & Herl, H. (in press). Hypertext as a strategy for teaching and assessing knowledge representation. In S. Dijkstra (Ed.), *Instructional models in computer-based learning environments*. Berlin: Springer-Verlag.
- California History-Social Sciences Framework*. (1988). Sacramento: California State Department of Education.
- Linn, R.L., Baker, E.L., & Dunbar, S.B. (1991). Complex, performance-based assessment: Expectations and validation criteria. *Educational Researcher*, 20(8), 15-21.

## Chapter 2

### Guidelines for Using CRESST's Model for Assessing Explanation

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

***CRESST's approach***

- a Prior Knowledge Measure which assesses (and activates) students' general and topic-relevant knowledge;
- provision of primary source materials (text), that is, new information in written text for students to read;
- a writing task in which students integrate prior and new knowledge to explain subject matter issues in response to a contextualized prompt;
- the scoring rubric for the writing task.

The rest of this section discusses the characteristics of these elements, particularly the rubric for scoring student writing, and the procedures for conducting the assessment. Sample tests are presented in Chapter 6.

#### Assessment Task Elements

**Prior Knowledge Measure.** This test consists of 20 short-answer items on concepts and facts related to the text material, to the general topic (e.g., historical period) in question, and to basic principles and concepts that recur in the content area (American historical thought). Typical items include “state’s rights” and the “Kansas-Nebraska Act” (history) and “density” and “acid” (chemistry).

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

***Concepts and facts related to the text material***

***Activates  
students' existing  
knowledge***

deep understanding and the ability to connect new information (from the primary-source texts) to ideas, concepts, and facts they have already learned.

Giving a prior knowledge test before having students read new text material activates students' 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.

**Text material.** The provision of primary source materials for students to read and incorporate in their writing is an important element of our assessment task. The text materials are excerpted from primary sources, typically speeches or letters for the history topics and descriptions of experiments for the science topics.

Students were given time to read two opposing views on the same topic, such as a Lincoln-Douglas debate, and their essays used these readings as a point of departure. In the development and validation of our materials, we also included short, multiple-choice reading comprehension measures after students read the text material. Those measures were helpful in guiding our selection and editing of primary texts during our test development phase, but we do not believe they are essential to the assessment process.

***Students'  
thinking and  
writing  
benefit from long  
composing  
periods***

**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.

***Issues using  
information in  
the text***

Our research suggests that students' thinking and writing benefit from long composing periods. 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.

**Example 2.1  
Writing Assignment Prompt**

Name \_\_\_\_\_

**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 Stephen 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.

---

***Contextual prompts to heighten student understanding of the task***

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, for example, “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, such as “You want to explain to your cousin....” Although we found no significant improvement from using such prompts, they may 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 CRESST essay scoring rubric.** We employ a criterion-referenced scoring strategy. Papers are judged in terms of pre-specified standards rather than in relationship to one another.

***Scoring rubric for the essay***

The scoring rubric for the essay consists of six scales or dimensions:

1. General Impression of Content Quality;
2. Prior Knowledge;
3. Number of Principles or Concepts;
4. Argumentation;
5. Text; and
6. Misconceptions.

*General Impression of Content Quality* rates 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).

*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.

**Scoring rubric  
for the essay**

*Number of Principles or Concepts* is a count of the major principles or concepts that students use with comprehension in their essay.

*Argumentation* is a measure of how well the student organizes knowledge to present a convincing argument or a coherent explanation.

*Text* (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 Chapter 6 of this handbook.

**Scales of the  
scoring rubric  
are differentially  
useful**

Each essay is scored on all six dimensions using a scale ranging from 0 to 5. Our studies indicate that, in general, the best essays receive low-to-moderate scores for *Text* (proportion of essay using text based detail), with high scores on the other scales. Note that a high score on *Misconceptions* indicates few or no misconceptions.

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 emphasized, then there may be a reason to look for high scores on the *Text* dimension. When modifying the scoring rubric, directions to students should be revised to match instructional goals.

**High levels of  
interrater  
reliability**

We have achieved high levels of interrater reliability for the scoring rubric across different topics and different raters (see Chapter 3). The lowest reliability estimates have been associated with the dimension of *Misconceptions* (correlation of .68).

***High levels of agreement can only be obtained***

Relatively high levels of agreement are essential for valid use of the scale. Such levels of agreement 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 (see Chapter 5).

### **Conducting the Assessment**

**Schedule of tasks.** In a typical (non-experimental) assessment situation, the approximated times required to administer the materials are as follows:

- 15 minutes to complete prior knowledge measure;
- 25 minutes to read primary source text (depending on students' reading ability);
- 50 minutes to read prompt and write essay.

Longer reading and writing tasks may be used instead of those provided here, as long as adaptations meet the specifications for task development discussed in Chapter 4. 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 large selection provided, 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 introduction and description of assessment tasks;
- 10 minutes to read and learn the instructions (script) for testing and familiarize oneself with the measures;
- 90 minutes to administer the student tasks (Prior Knowledge Measure, reading, and essay).

***Basic tasks and schedule***

***Reassure  
teachers***

We asked teachers to complete the Prior Knowledge Measure themselves in order to provide us with 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 assessments on new topics. 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 assessment packets.** Packets of student materials should be organized ahead of time, with students’ names, ID numbers, teacher, school, and prompt version typed on a label placed on the front of their packets. This will facilitate management of packets during the assessment period. Color-coded materials also simplify administration procedures.

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 Measure and essay page 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 introduction 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
- samples of materials given to students.

Teachers should be given these materials well before the time of assessment so that they may become familiar with the procedures and purposes and have time to resolve questions.

***Packet of  
materials***



***Sample  
assessments***

## Chapter 3

### Sample Assessment Materials for Students (U.S. History and Chemistry Topics)

This section of the handbook contains two sample sets of student assessments: the Civil War era in U.S. history and analysis of an unknown substance (chemistry). You may duplicate and use these materials as presented here or you may use them as models for developing your own assessment program.

**Example 3.1  
Prior Knowledge Measure  
History**

Name \_\_\_\_\_

#### 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 \_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

**Example 3.2**  
**History Assessment Texts**

Name \_\_\_\_\_

**Texts: Lincoln-Douglas Debate\***

**Directions:** As Abraham Lincoln and Stephen 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,

***History Assessment Texts (continued)***

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 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

***History Assessment Texts (continued)***

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 electional 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.

***History Assessment Texts (continued)*****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

***History Assessment Texts (continued)***

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 inwith 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



***History Assessment Texts (continued)***

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 unequalled 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.

\* From *Political Debates Between Abraham Lincoln and Stephen A. Douglas* (Cleveland, 1902), pp. 43-47.

**Example 3.3**  
**History Writing**  
**Prompt**

*The following  
prompt is based  
on the texts  
provided in  
Example 3.2*

Name \_\_\_\_\_

**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 Stephen 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.

---

**Example 3.4**  
**Prior Knowledge Measure**  
**Chemistry**

Name \_\_\_\_\_

**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 as it relates to chemistry 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 \_\_\_\_\_

\_\_\_\_\_

**Example 3.5**  
**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. She then added a portion of soda A to one test tube and an equal portion of soda B to another test tube, making sure to note on each test tube which soda was added. The third test tube 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 test tubes containing soda 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 A and 100 ml of soda B 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 prepared new samples containing equal amounts of each soda and added equal amounts of salt to each sample. 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 soda 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:

	A	B
Yeast test	distinct odor gas bubbles	no odor 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 dissolved
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.”

*\*This task was adapted with permission from one developed and tested by the Connecticut State Department of Education.*

**Example 3.6**  
**Chemistry Writing**  
**Prompt**

Name \_\_\_\_\_

**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.



# Chapter 4

## Specifications for Developing Assessment Materials

***Do supposedly similar assessment tasks tap the same intellectual processes?***

***A common template or framework must guide the development of such tasks***

The demand for alternative ways to assess students' complex thinking skills rather than mere recall of facts has resulted in a dramatic increase in the development of new measures for use at the local or 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.

***Parallel  
assessment tasks***

***Specifications to  
control  
structure and  
content***

### **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.

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 4.1. The specifications given here are specific to our history assessment and can easily be adapted to other content areas or different local constraints.

**Table 4.1  
Assessment Components  
Requiring Specification**

1. Text Materials
2. Prior Knowledge Measure
3. Essay Task



### **Specifications for History Text Materials**

We developed the set of specifications to cue the selection of text for use in assessing history. Please see the next page.

**Table 4.2**  
**Specifications for History Text**  
**Materials**

Topic	<ul style="list-style-type: none"> <li>• Must be a regular and significant part of the secondary school history curriculum.</li> <li>• Must provide an issue that has implications beyond the particular historical period.</li> </ul>
Text Structure and Form	<ul style="list-style-type: none"> <li>• Must provide for contrasting views, explanations or contexts.</li> <li>• May use either a single piece or short contrasting pieces.</li> <li>• May be written in narrative or expository form.</li> <li>• Should be short enough to read within a class period or less.</li> <li>• Must be written so that esoteric or technical discussions are minimal, special vocabulary is limited, and the author's point of view is clear.</li> </ul>
Text Source	<ul style="list-style-type: none"> <li>• Must use primary source materials.</li> <li>• Accepted pieces include letters, transcripts of speeches, editorials, excerpts from documents.</li> <li>• As a rule, provide only two major text selections. However, supplementary materials, including written personal reactions by historical figures, maps, songs, and other relevant material may be made available, so long as adequate time is provided for the students to understand the materials. You may tell the students that these materials are "required" or "optional."</li> <li>• Editing of materials is to be avoided except to excerpt sections from a longer piece.</li> </ul>

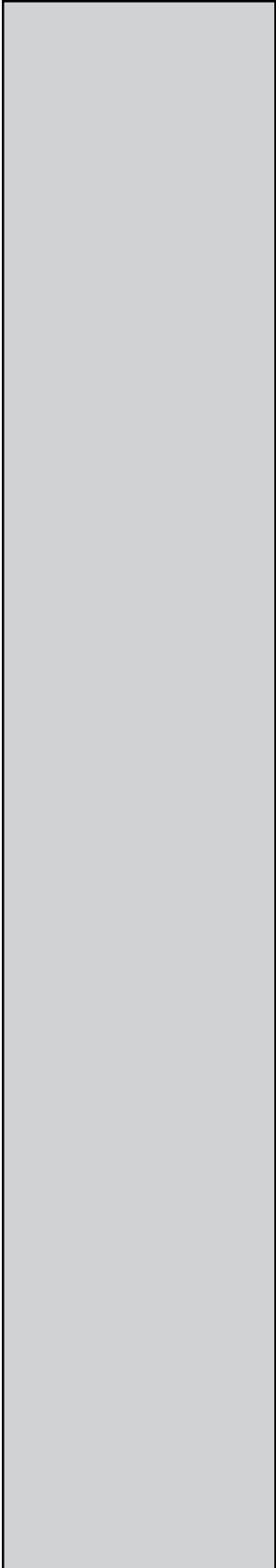
***General  
assessment of  
students'  
knowledge of  
American  
history***

**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 are 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 4.3.

**Table 4.3**  
**Specifications for the Prior Knowledge Measure**

Stimulus Format	<ul style="list-style-type: none"> <li>• Proper names, terms, numbers, and short sentences not to exceed 8 words.</li> </ul>
Forms of Knowledge	<ul style="list-style-type: none"> <li>• Specific facts, events, dates, quotations, the names of principles or concepts.</li> </ul>
Distribution	<ul style="list-style-type: none"> <li>• Half the items should reference specific information, such as an event (Harper's Ferry Incident) and half of the items should reference concepts (states' rights) or principles (constitutionality).</li> <li>• At least two-thirds of the items should be relevant to the immediate historic period of the assessment (plus or minus 10 years of the date(s) of the text). The remaining one-third of the items can precede or follow the period under assessment.</li> </ul>
Directions to Students	<ul style="list-style-type: none"> <li>• Students should be encouraged to respond rapidly and to write the essence of their understanding briefly. There are no requirements for form, for example, complete sentences.</li> </ul>
Administration Constraints	<ul style="list-style-type: none"> <li>• Approximately 20 items can be provided in a 10-15 minute period.</li> </ul>
Scoring Scheme	<ul style="list-style-type: none"> <li>• Responses are scored on a five-point scale (0=low; 4=high). Students are given a "4" if they have an accurate, elaborated definition, description or context for the stimulus term; a "3" if they are essentially correct; a "2" is assigned if they have some incomplete notion of the term; a "1" if they have no idea; and a "0" if they make no response.</li> </ul>



Specifications for the essay task 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 4.4 on the next page.

**Table 4.4**  
**Essay Task Specifications**

Essay Context	<ul style="list-style-type: none"> <li>• Students should be given an historical context to frame the written response, consisting at minimum of a time, historical period, and occupational role.</li> </ul>
Audience	<ul style="list-style-type: none"> <li>• The audience for the student's writing is specified to be a particular person in the same target historical period. The person must be ignorant of the information provided in the texts for some plausible reason, such as living abroad or returning from a long trip, to heighten the verisimilitude of the task.</li> </ul>
Intellectual Task	<ul style="list-style-type: none"> <li>• 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, to compare and contrast perspectives using inference strategies, and to synthesize the explanation referring both to relevant text material and prior knowledge.</li> </ul>
Directions to the Student	<ul style="list-style-type: none"> <li>• 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.</li> </ul>
Administration	<ul style="list-style-type: none"> <li>• 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.</li> <li>• Students should have the text selections available to them as they write.</li> <li>• 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.</li> </ul>
Scoring Scheme	<ul style="list-style-type: none"> <li>• Essays should be scored in terms of the six scoring dimensions, General Impression of Content Quality, Prior Knowledge, Number of Principles or Concepts, Argumentation, Text, and Misconceptions, as described at length in the training materials.</li> </ul>



***Tasks for periods  
of interest***

Following the use of specifications to generate comparable assessment tasks for periods of interest, for instance, 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.

# Chapter 5

## Rater Training, Scoring and Reporting

### *Essay rating procedure*

This section provides guidelines for organizing and conducting rater training and scoring sessions. It uses the CRESST essay scoring rubric and prior knowledge scoring scale. This section also addresses reporting options. We emphasize the essay rating procedure because of its importance and because this procedure is more difficult than scoring the prior knowledge measure.

### **Who Can Rate?**

We have found that people with strong content area backgrounds, acquired either as undergraduates or through later course work, make particularly good raters. We have successfully trained raters with widely varying backgrounds, including elementary and secondary school teachers, substitute teachers, graduate students, subject-matter experts, and district office administrators. Since essay scoring in a content area 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.

### **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 per day before losing attention and precision. We have also found that when possible, it is better to use more raters for a shorter period of time to maintain a certain “momentum” to the rating.

Based on a 75-papers-per-reader-per-day guideline, divide the total number of papers you will be scoring by 75. Divide this number by the number of days you have available for scoring to determine the actual number of raters needed.

*More raters for a shorter period of time*

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. This will tell you 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:** ..... 3,000 papers to score

**Guideline:** ..... 75 papers per day per person (6 hour day)

**Rater Days:** .....  $3000/75 = 40$  rater days

**Guideline:** ..... More raters over fewer days, if possible

**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 (to be described later).

When possible, train slightly more raters than you think you will need. There may be attrition due to illness, or you may find that some raters cannot be brought up to an acceptable level 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). This is because he or she will need to become familiar with the primary text materials for each prompt and practice applying each scale to writing on different topics.

***Train more  
raters than you  
think you'll need***

***Recruit good  
raters***

***Limit actual training and rating time***

You will need to schedule scoring so that you have the requisite three to four 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 six hours, not including breaks and lunch. Schedule training with a brief informal period of 10 minutes at the start. This way raters can meet and late arrivals will be less disruptive.

In our experience, we have found that six to eight raters is the maximum number that one person can effectively train. When the number of raters is larger than this, the raters should be divided into groups of six to eight that will work at separate tables or in separate rooms. Each group should have its own trainer or table leader.

**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 and 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

***Sample schedule***

***Schedule for  
training and  
scoring***

**Sample Schedule for Training and Scoring Sessions**

**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)

***Trainer  
responsibilities***

**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 leader 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 devise the "set-up" for the training session, assignment of raters to different scoring tables, and facilitate luncheon arrangements;

***Trainer  
responsibilities***

4. to set up the training and scoring schedule and select table leaders;
5. to read the model essays, training essays, criterion essays, and check essays and be able to explain how scores match the rubrics for every topic;
6. to organize the papers to be scored, have papers coded if necessary, organize into sets for raters, and bring to the scoring site;
7. to organize the reliability check at the end of training and schedule the check papers to maintain high rater reliability;
8. to retrain, dismiss or reassign “unreliable” raters; and
9. to supervise the reproduction of training materials (prompts, rubrics, training papers, model essays, essay scoring summaries) and organization into rater packets.

*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 (six to eight per table including the table leader). Ideally, you should have an overhead projector to project sample papers during rater training. Set aside space for “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 copy, distribute, and discuss unanticipated problems arising from any papers that differ dramatically from those in the training materials.

***Room set up***

***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 materials:

**Table 5.1**  
**Materials Necessary for Training**

<b>Materials</b>	<b>Users</b>	
	<b>Raters</b>	<b>Trainer</b>
<b>1. Scoring Guidelines</b>	<b>x</b>	<b>x</b>
<b>2. Scoring Recording Forms</b>	<b>x</b>	<b>x</b>
<b>3. Prescored, Typed Model Essays</b>	<b>x</b>	<b>x</b>
<b>4. Unscored, Typed Model Essays</b>	<b>x</b>	
<b>5. Prescored Training Essays - handwritten</b>		<b>x</b>
<b>6. Unscored Training Essays - handwritten</b>	<b>x</b>	
<b>7. Prescored Criterion Essays</b>		<b>x</b>
<b>8. Unscored Criterion Essays</b>	<b>x</b>	

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 on the next page.

***Time for reproduction of training materials***

## **Suggested Organization of Rater Training Materials**

(Notebook with Dividers)

### **Rater training materials**

- Section I:** Copy of Prior Knowledge Measure, primary text materials given to students, and actual writing prompt. (If you are scoring papers for several prompts, keep these in separate sections.)
- Section II:** Copy of the scoring rubric and guidelines
- Section III:** Score Recording Forms
- Section IV:** Training Examples to illustrate each scale (typed)
- Section V:** Prescored and Unscored Training Essays (handwritten)
- Section VI:** Criterion Essays (handwritten)

*(Create the same type of rater training materials for each separate topic.)*

We estimate that a minimum of 20 papers need to be available for training in a single task and approximately three to four more for each additional topic area. Chapter 6 contains examples of such papers for training purposes. You may duplicate these papers. The conceptual background in Chapter 1 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. Raters (teachers) might describe 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.

### ***Raters introduce themselves***



***Answer questions  
about the tasks***

**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 should provide an overview of the six scales. When questions about the rubric as a whole have been answered, training should move on to the specifics of each scale—using typed, prescored training examples 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 *Text* scale. Raters may find it useful to list some of the concepts or principles that may frequently occur in the essays—facilitating initial application of the *Number of Principles or Concepts* scale.

***Provide raters  
with practice***

Eventually the trainer can move to an unscored essay, letting 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 a new scale until the group understands the current scale well. After all six 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 should 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. Afterwards, the trainer will provide corrective feedback using the prescored versions.

In applying the full rubric, ask the rater to read the paper and always score *first* the *General Impression of Content Quality* scale. The rater should then 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 in the real scoring session will be provided in typed form (such as those completed on a word processor) then the portion of training using handwritten student papers could be skipped.

**Rater reliability assessment.** The last training task is to assure that raters:

- are assigning scores on the basis of the rubric;
- agree with each other in the assignment of scores; and
- rate at a reasonable speed (about 2-5 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

***Feedback on the correct scores***

***Reliability assessment***

***Use rubric  
guidelines***

*not* model essays, training essays, or practice essays. They are a special set of three to five essays, prescored by experts, to be used at the end of the training period. These essays will be used to determine if raters are ready to rate actual student work.

At this time, 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. The packet 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 conduct the reliability check.

***Rater reliability***

To calculate rater reliability on the criterion papers, follow these steps and see the example in Table 5.2:

1. Create a table like Table 5.2, highlighting the expert's rating. If plus-or-minus ( $\pm$ ) 1 agreement is acceptable, highlight the range of acceptable scores 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 *General Impression of Content Quality* scale, which was the "expert" rating.
3. Count the number of raters who gave the paper the expert rating. These are the raters with "perfect agreement." Enter the number in the appropriate column.

***Rater reliability***

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 ( $\pm$ ) 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” results and the “Percent +1 Agreement” results in the Averaging Table that follows.
9. Calculate the average for each of the two kinds of percent agreement for each scale. In Table 5.3, 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 “( $\pm$ ) 1 Agreement” for that scale is 90%.

**Table 5.2**  
*example of*  
**Calculating Reliability**  
**Based on 2 Criterion Papers**  
 (Total # of Raters = 10)

Key  
 criterion rating  
 criterion  $\pm 1$

Paper #1 Scales	Frequency Count of Raters' Scores						# of Ratings That Match Criterion Ratings	% Perfect Agreement	# of C $\pm$ 1 Ratings	% $\pm$ 1 Agreement
	0	1	2	3	4	5				
GICQ		II	III	II	I		5	50%	9	90%
PK		III I	II		I	I	6	60%	8	80%
PN	I	III II	II				7	70%	10	100%
A			I	III III	I		8	80%	10	100%
TEXT					III III		10	100%	10	100%
MIS				II	II	III I	6	60%	8	80%
Paper #2										
GICQ			I		III III	I	8	80%	9	90%
PK			I		III III		8	90%	9	90%
PN			I	III III	I		9	80%	10	100%
A			I	III II		II	7	70%	8	80%
TEXT		III III	II				8	80%	10	100%
MIS			I	I	III II	I	7	70	9	90%

*Note:* GICQ = General Impression of Content Quality, NP = Number of Principles or Concepts, PK = Prior Knowledge, A = Argumentation, TEXT = Text (proportion of essay using text-based detail), MIS = Misconceptions

**Table 5.3**  
**Averaging Table**  
**% of Rater Agreement**

<b>Averaging Table</b>						
Scales	% Perfect Agreement			% $\pm 1$ Agreement		
	Paper #1	Paper #2	Average	Paper #1	Paper #2	Average
G I C Q	50%	80%	65%	90%	90%	90%
N P	70	80	75	100	100	100
P K	60	90	75	80	90	85
A	70	80	75	100	80	90
T E X T	100	80	90	100	100	100
M I S	60	70	65	80	90	85

*Note:* GICQ = General Impression of Content Quality, NP = Number of Principles or Concepts, PK = Prior Knowledge, A = Argumentation, TEXT = Text (proportion of essay using text-based detail), MIS = Misconceptions

***Criterion score  
for each scale***

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  $\pm 1$  Agreement should be even higher (e.g., 90-100%). (This is because the latter is actually quite a liberal difference from the criterion score on a scale with only six score 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.

***Benefit of a  
larger number  
of raters***

If a small group of raters is inconsistent with the rest of the group, you may wish to re-train 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 Measure for the same tasks (directions follow). It is our experience that scoring the Prior Knowledge Measure is a much simpler procedure.
2. Re-train the rater. This option 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 re-training can occur after the rest of the group starts rating the actual essays.

***Raters who  
consistently  
tend to rate  
inaccurately***

If low scores would result in significant consequences for the 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.

***Another  
reliability  
assessment***

As long as most of the raters fail to agree, you will need to continue training and then conduct another reliability assessment before proceeding to the scoring session. If this is not successful, the following options are available:

1. Decide to go with a long rating period using those few raters who do agree.
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 (usually 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 5.4 lists the materials you will need for the actual scoring session. Details about preparing them are discussed below.

**Table 5.4  
Materials Required for the  
Scoring of Local Essays**

***Materials  
you need***

Materials	Users	
	Raters	Trainer
1. Packets of Student Essays with interpolated Check Papers	<b>x</b>	
2. Prescored Check Papers		<b>x</b>
3. Scoring Guidelines	<b>x</b>	<b>x</b>
4. Scoring Sheets	<b>x</b>	<b>x</b>
5. Calculator		<b>x</b>



***Reliable measure  
of a student's  
writing***

**Single or multiple scores per paper.** Prior to the scoring session, you must decide whether each student 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 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.

***Assigning scores  
consistently***

You can then calculate rater agreement much as was done to assure reliability during rater training or using more sophisticated statistical techniques. This will help 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 of essays in a different random order.

***Rater pairs  
randomly  
assigned***

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. For example, each paper in the 20% sample is read by two raters rather than by all of them. 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

***Multiple scoring  
of each paper***

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*, we advise multiple scoring of *each* paper. 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.

***Monitor scoring***

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 sets of papers.** Once you have developed your rating plan (e.g., single scores for 80% of the essays and double scores for a 20% sample, double scores on marginal papers, or multiple scores for all), you will need to assign raters their initial sets of papers. You also need to 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.

***Randomly  
ordered***

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.

***Sort by topic and  
grade level***

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 or topic, 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 1,000 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 to 20.

***Order of prompts  
across raters***

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

***Machine-  
readable  
document***

papers. You should have at least four 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.

**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 so often provides a system for monitoring rater reliability. This is an especially difficult task when the number of raters exceeds 20 or when rating is done at different sites and on 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.

***Dependent upon  
quality of rater  
training***

***One check paper  
per hour***

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 raters to pull their scoring sheets and the check papers 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”.

***Retraining***

The trainer should calculate rater agreement by 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 them in a small group. If, over a series of check papers a previously qualified rater begins to lose his or her reliability, 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 writing task assessment results is to compare an individual's or group's average scores on the six scales to profiles of the typical expert or novice. As Figure 5.1 indicates, a content *expert* tends to score high on *General Impression of Content Quality*, *Prior Knowledge*, *Number of Principles or Concepts*, *Argumentation*, and *Misconceptions*, but moderate-to-low on the *Text* scale.

***Rely heavily on  
the text***

A content *novice* tends to do the opposite, to exhibit little conceptual or prior knowledge and to rely heavily on the text. A few novices have several misconceptions and display them in their essays. Other novices score well on the *Misconceptions* scale either because they truly have few misconceptions or because they are aware that their knowledge is shaky and manage to avoid revealing the misconceptions in their essays.

**Figure 5.1**  
**Expert and Novice Profile Tendencies**

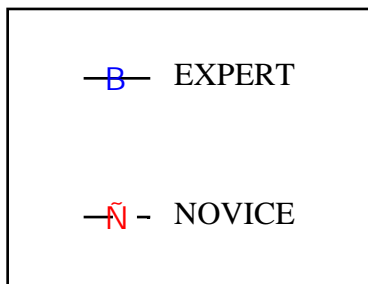
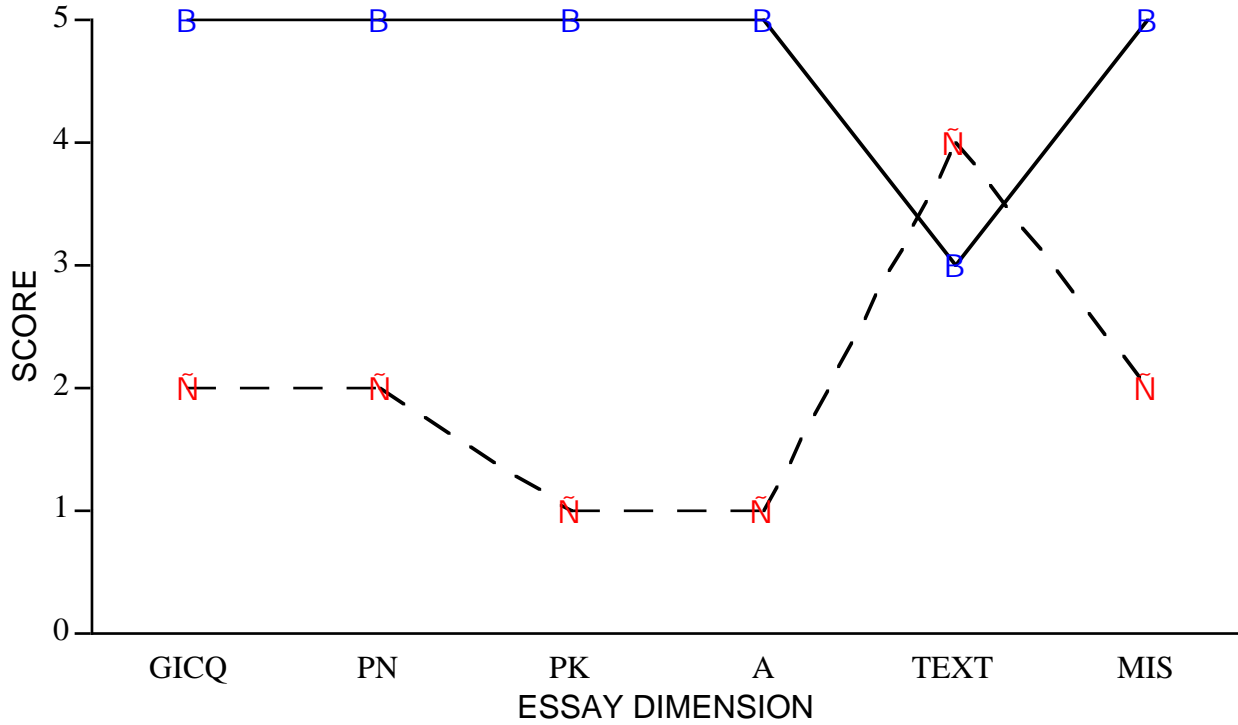


Figure 5.1. Expert and novice profile tendencies. GICQ = General Impression of Content Quality, PN = Number of Principles or Concepts, PK = Prior Knowledge, A = Argumentation, TEXT = Text (proportion of essay using text-based detail), MIS = Misconceptions.

***Qualities  
of  
content expertise***

Scores for two high school students' essays are depicted in Figure 5.2. Student #1's profile shows qualities associated with content expertise, as discussed above. Contrast this with the profile for Student #2, with a relatively high *Text* score, a minor misconception (score point 4), and low scores on the other dimensions. This is a profile more typical of a novice level of understanding.

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 essays) 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.

***Effects of  
variables***

**Scoring and Using the Prior Knowledge Measure**

Scoring the Prior Knowledge Measure 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 the student's responses. Here, writing quality and grammar definitely do *not* count. Scoring guidelines and model tests may be found in Chapter 6 for your use. In our experience, it takes an average of 1-2 minutes to score a test, depending on the student's level of knowledge.

**Figure 5.2**  
**Sample Student Profiles**

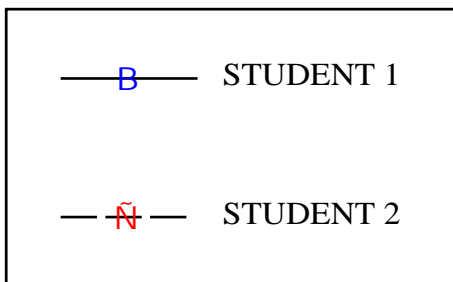
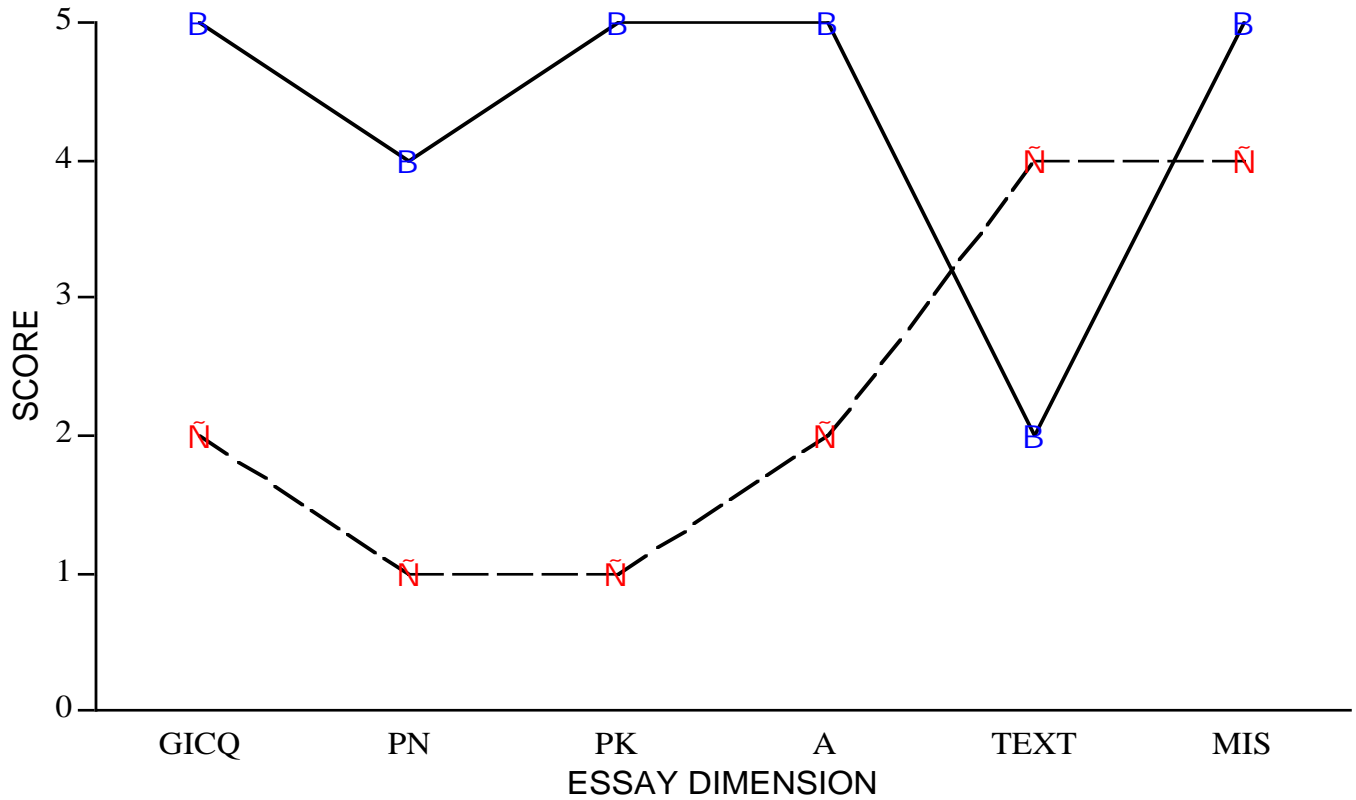


Figure 5.2. Sample student profiles for Civil War essay topic. GICQ = General Impression of Content Quality, PN = Number of Principles or Concepts, PK = Prior Knowledge, A = Argumentation, TEXT = Text (proportion of essay using text-based detail), MIS = Misconceptions.



Generally, raters can be trained rather rapidly to high levels of agreement (95-100%) by using the rubric in Chapter 6.

When checking rater agreement, plus-or-minus-one-agreement may be skipped as too liberal for this 5-point scale. It is usually sufficient to double-score only a 20% sample of the Prior Knowledge Tests.

***Double score only  
a 20% sample***


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 Measure can be reported and used in several ways, depending on your purpose. In one approach, 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 an important 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.

***Prior knowledge  
items***

Another way of using the findings to improve instruction is to group the prior knowledge items into subsets (e.g., facts and concepts) 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 of activating



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.

# Chapter 6

## Sample Training Materials

### *Training materials*

This section contains the following materials:

- A. Prior Knowledge Scoring Guidelines
- B. Scored Prior Knowledge Measure
- C. Essay Scoring Guidelines
- D. Essay Scoring Rubric
- E. Training Examples #1-10: Student essays scored and annotated to illustrate individual scales (typed)
- F. Training Essays #1-3: Student essays scored on all scales (handwritten)

### **A. Prior Knowledge Scoring Guidelines**

### *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

**Guidelines**

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 mak-  
ing things by machines. Contributes to growth of  
cities/source of difference between north and south.

B. an elaborated definition that indicates complete under-  
standing or knowledge of term.

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

Errors should be ignored and the part of the answer that is  
correct should be scored.

<b>B. Scored PK Measure</b>
-----------------------------

Name: \_\_\_\_\_

**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.

0 1. popular sovereignty \_\_\_\_\_  
 \_\_\_\_\_  
 \_\_\_\_\_

- 1 2. Dred Scott It was a case of Supreme Court dealing with a white man's slave & the white man
- 1 3. Communism form of government that doesn't give people freedom. ex. Russia
- 2 4. Missouri Compromise So that the # of states would equal each other between slavery & non-slavery
- 4 5. industrialization the coming of machines, factories business 1920's
- 3 6. Gold Rush In California- gold was found, thousands rushed to try & get rich
- 0 7. bleeding Kansas \_\_\_\_\_
- 1 8. states' rights the state is allowed to have laws so it can govern
- 0 9. Federalism \_\_\_\_\_
- 2 10. underground railroad to try & help slaves get to the north
- 1 11. imperialism part of government
- 2 12. Whig party of the government didn't last long
- 0 13. Kansas-Nebraska Act A land act that said you could buy so much land there was a minimum & so much would be set aside for public use

- 2 14. Abolitionists they wanted slavery outlawed
- 0 15. sectionalism \_\_\_\_\_
- 3 16. westward movement people moved out west because land was cheap & more of it
- 0 17. constitutionality \_\_\_\_\_
- 0 18. New Deal \_\_\_\_\_
- 0 19. party platform where a party can get up & speak its views
- 0 20. balance of power During WWI this was disrupted because Italy became a country & 1 cause of the war.

*Essay scoring  
guidelines*

## C. Essay Scoring Guidelines

### 1. **General Impression of 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. **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 essays. 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; for example, “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

Example: At Harper’s Ferry, John Brown attempted to lead a slave revolt but failed.



***Essay scoring  
guidelines***

**3. Number of Principles or Concepts (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 of 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.”

***Essay scoring  
guidelines***

Score point guidelines:

0 - no response

1 - no concepts/principles

2 - one concept/principle

3 - two concepts/principles

4 - three concepts/principles

5 - four or more concepts/principles

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 economy and coasts. Also, with Britain there was a great advantage with exportation. It seemed our economy could only suffer without the aid of Britain."

**4. Argumentation (A)**

How well does the student organize historical knowledge to make a convincing argument? (0-5 point global rating: 0 = no response, 5 = highest level of coherent and cohesive argumentation)

This scale focuses on how well the student analyzes and organizes historical evidence to make a well-reasoned argument or interpretation. Essays scoring at the highest level will provide adequate evidence to support and justify interpretive stances and a chain of logical argumentation or analysis. In the development of plausible analysis or agreement, coherence will naturally be important. A paper that lacks coherence or logical flow should not earn the highest score.

**5. 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.

***Coherence***

***Essay scoring  
guidelines***

Assign score according to point which comes closest to the proportion of text detail in the student's essay. For instance, in a four-to-five page essay, 1 or 2 sentences reflecting information from the provided text will earn a TEXT score of 1. In this case, the proportion of material from the text is closer to 0 than to 1/4.

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

**6. Misconceptions (MIS)**

This is a measure of the amount of incorrect information, or the number of misconceptions or misinterpretations, in the essay. Note that a high score on Misconceptions indicates few or no misconceptions.

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 - one or more serious misconceptions central to the essay

2 - at least one serious misconception

3 - several minor errors an/or a moderate misconception

4 - very minor misconception

5 - no misconceptions

***Incorrect  
information***

***Wrong opinions***

*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.

## D. Essay Scoring Rubric

*Content quality*

### 1. General Impression of 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)

*Facts and events*

### 2. 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

### 3. Number of Principles or Concepts (PN)

0 - no response

1 - no principles/concepts

2 - one principle/concept

3 - two principles/concepts

4 - three principles/concepts

5 - four or more principles/concepts

*Organize  
historical  
knowledge*

### 4. Argumentation (A)

How well does the student organize historical knowledge to make a convincing argument? (0 - 5 point global rating: 0 = no response, 5 = highest level of coherent and cohesive argumentation)

***TEXT*****5. 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

***MIS*****6. Misconceptions (MIS)**

0 - no response

1 - one or more serious misconceptions central to the essay

2 - at least one serious misconception

3 - several minor errors an/or a moderate misconception

4 - very minor misconception

5 - no misconceptions

***Training  
examples*****E. & F. Scored Student Essay Training Examples  
and Training Essays****Key to Scale Abbreviations**

GICQ: General Impression of Content Quality

PK: Prior Knowledge: Facts and Events

PN: Number of Principles or Concepts

A: Argumentation

TEXT: Text Proportion of Essay Using Text-Based Detail

MIS: Misconceptions

*Note: Where appropriate, content relevant to the score is underlined in Examples 1-10.*

***Training  
examples*****E. #1-10: Student Essays**

Scored and annotated to illustrate single scale (Typed)

**Training Example 1**

**Topic: Civil War**

**Scale: GICQ**

**Score: 4**

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 federacy 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

***Training  
examples***

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.

**Training Example 2**

**Topic: Revolutionary War**

**Scale: GICQ**

**Score: 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 feet as though History going to be kinder for me because I don’t enjoy History, but I will give it a shot.

*Training  
examples*

**Training Example 3**

**Topic: Revolutionary War**

**Scale: PN**

**Score: 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 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



***Training  
examples***

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 war, 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.

**Training Example 4**

**Topic: Civil War**

**Scale: PK**

**Score: 5**

It is the year 1858, and all is not well in the United States. Many years of bitter rivalry between the North and the South have manufactured into deep hatred for one another. The differences between the two do not result strictly from their views on the slavery issue, but on the differences in their economies, their views on the nature of the Federal Union, their ideas on who should control the central government, and finally their differences in civilization.

Their economic differences stem from the geographical conditions in which they live. It is a struggle between the industrial North and agricultural South. Bitter sectional rivalry on such issues as the protective tariff, the United States Bank, and slavery, have resulted in many bloody and meaningless conflicts. Slavery was obviously the most sensitive of all the issues, since it denied human freedom and isolated democratic ideals.

***Training  
examples***

The South insisted that the Federal Union was a pact among the states, that gave any state the right to secede. The North insisted that the Union was created by the people as one “indivisible nation” and that no state had the right to secede. Lincoln himself asserted that if there was to be a war, it would be to preserve the Union, and not to abolish slavery.

Both the North and South realized that control of the central government would pass to the section that gained the West’s support. Therefore, the South favored, and the North opposed the extension of slavery into the new territories. With time Western interests identified with those of the North’s railroads and industrial manufacturing. Thus, it became obvious that the South was losing its plea for Western support, as well as the fact that they still remained a minority in the Union.

The differences in civilization between the two, arose from two basic issues. One, the civilization in the South was geared more to the land owning aristocracy and the plantation. In the North the ideals were more democratic and dynamic. Two, the South lived in an area of sectional loyalty, whereas the North triumphed with national loyalty.

These issues became the platform for Lincoln and Douglas’s’ debates, along with other prominent issues such as popular sovereignty, the Dred Scott vs. Lanford case, the Missouri Compromise, bleeding Kansas, a horrible result of the Kansas-Nebraska Act of 1854, states’ rights, to secede or not secede?, and most important regarding the two sections of the country, sectionalism.

The issues presented will never be agreed on by Lincoln or Douglas, the North or the South, the Democrats or the New Present Day Republicans, or anyone on opposite sides of the 36’30 parallel. The issues stemmed from many years of discussion and careless thinking. Civil Rights were practically thrown out the door, and Civil War will soon be the horrible result.

*Training  
examples*

**Training Example 5**

**Topic: Civil War**

**Scale: TEXT**

**Score: 5**

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 Stephen 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.

**Training Example 6**

**Topic: Revolutionary War**

**Scale: TEXT**

**Score: 1**

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

***Training  
examples***

so that others could not leap 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.

**Training Example 7**

**Topic: Revolutionary War**

**Scale: MIS**

**Score: 1**

The American Revolution's colonies have just declared their independence from Great Britain. 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. Alot 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 Britain 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

***Training  
examples***

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. There is not a way to live. After all the battles in between only the lucky ones did live. Finally, they were free.

**Training Example 8****Topic: Civil War****Scale: MIS****Score: 4**

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 tritios. The abolitionists 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 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.

*Training  
examples*

**Training Example 9**

**Topic: Civil War**

**Scale: A**

**Score: 1**

Wow-EE Boy, I jus got back from that-der Lincoln-Douglas guvment d'bate. An Lemee tel yall it was the most intrestin d'bate with awl these new fangled idear's I jus dont know where ta start. First off I aught ta tell ya that der are many problems dat face dis here nation of ours. One of the problems is slavery. I don't reckon I know what the problem is so I figur dat both Linclon and Douglas is both sorta right in there figuring.

**Training Example 10**

**Topic: General Immigration**

**Scale: A**

**Score: 4**

Immigration... the issues on it are very strong and very complex. The debate of Julian Simon, a professor of sociology, believes that bringing in immigrants can only benefit the people. Otis Graham Jr believes however that, there are too many people for the standards of life that we wish to uphold. I believe that immigration is necessary because the immigrants have benefitted us in more ways than one.

"...average immigrant family takes less in welfare & pays more taxes than the average native family." This statement by Simon clearly states that the immigrants pay their share of taxes, but take less of the public services offered to them. Therefore helping the country economically by giving more \$ than they are taking. Simon believes that the immigrants also have a more hopeful, forward-looking, outlook on the economic future of America. He is knowledgeable to the fact that they may impose costs upon us, but in the long run can only benefit us by coming to America.

Graham has opposing views to immigration. He believes that "there are too many Americans already, for the margins of environ-

***Training  
examples***

mental safety & the standards of life which we wish to secure.” Graham also believes that the immigrants are squeezing out our resources & what we have left of our surrounding environment. The population, Graham states, consists of thousands of immigrants, & if we allow the #'s to rise, there may be an overpopulation of people in America.

I believe that it is important to have immigrants in the country. Indeed, to save our resources the #'s must be closely watched, but the immigrants have helped our nation by providing the taxes they pay & many have even invented things we Americans would never think of doing. Immigrants are hard-working people & given a chance could be very useful. Having been a generation of immigrants, I feel very fortunate to have had the chances I have had living here. I know others would feel the same. It is necessary not only for them, but for us as Americans.

Some may agree, and yet, some may disagree with immigration, but what people need to realize is that at one time, we were all immigrants. We came to America, “the land of opportunity” in hopes for establishing & reaching our goals and dreams. Many want to achieve that dream...only if they were given the chance to do so.

Scale:	GCIQ	PK	PN	A	TEXT	MIS
Score:	2	1	2	2	3	5

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 for 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 & sovereignty of each state and said that it was not possible to "establish uniformity in the <sup>local & domestic</sup> institutions of all the states..." ~~was~~ ~~and~~ ~~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.



Scale:	GCIQ	PK	PN	A	TEXT	MIS
Score:	2	1	1	2	4	4

On important issues such as slavery<sup>†</sup> and saving the Union, Lincoln and Douglas strongly disagree. In Douglas' speech, he didn't really talk about what he was going to do to save the Union, but more on disagreeing w/ everything that Lincoln plans on doing. On the other hand, Lincoln defends himself, saying that ~~being~~ Douglas misinterpreted the point that Lincoln was trying to get across.

The one basic thing that Lincoln and Douglas were debating about, ~~was~~ was slavery. Lincoln believes that the Union

cannot be preserved as half slave and half free Union. Douglas came back w/ the fact that it has stayed this way for 82 years. Lincoln is strongly against slavery in the Union whereas he says that the Union must be all slave or all free. Douglas isn't against slavery, but he also isn't proslavery.

The ~~Lincoln-Douglas~~ Lincoln-Douglas Debate was very interest~~ing~~ for the both of them since they were so against each other. It seemed like they were actually debating if the North & South were to go to war.

Scale:	GCIQ	PK	PN	A	TEXT	MIS
Score:	5	5	4	5	2	5

It is the year 1858, and all is not well in the United States. Many years of bitter rivalry between the North and the South have manufactured into deep hatred for one another. The differences between the two do not result strictly from their views on the slavery issue, but on the differences in their economies, their views on the nature of the Federal Union, their ideas on who should control the central government, and finally their differences in civilization.

Their economic differences stem from the geographical conditions in which they live. It is a struggle between the industrial North and agricultural South. Bitter sectional rivalry on such issues as the

2

protective tariff, the United States Bank, and slavery, have resulted in many bloody and meaning less conflicts. Slavery was obviously the most sensitive of all the issues, since it denied human freedom and violated democratic ideals.

The South insisted that the Federal Union was a pact among the states, that gave any state the right to secede. The North insisted that the Union was created by the people as one "indivisible nation" and that no state had the right to secede. Lincoln himself asserted that if there was to be war, it would be to preserve the Union, and not to abolish slavery.

Both the North and South realized that

3

control of the central government would pass to the section that gained the West's support. Therefore, the South favored, and the North opposed the extension of slavery into the new territories. With time Western interests identified with those of the North's railroads and industrial manufacturing. Thus, it became obvious that the South was losing its plea for Western support, as well as the fact that they still remained a minority in the Union.

The differences in civilization between the two, arose from two basic issues. One the civilization in the South was geared more to the land owning aristocracy and the plantation. In the North, the ideals were more democratic and dynamic. Two,

4  
the South lived in an area of sectional loyalty,  
whereas the North triumphed with national loyalty

These issues became the platform for  
Lincoln and Douglass' debates, along with other  
prominent issues such as popular sovereignty,  
the Dred Scott vs. Sandford case, the Missouri  
Compromise, Bleeding Kansas, a horrible result  
of the Kansas-Nebraska Act of 1854, states' rights,  
to secede or not secede?, and most important,  
regarding the two sections of the country, sectionalism

The issues presented will never be agreed on  
by Lincoln or Douglass, the North or the South,  
the Democrats or the New Present Day Republicans,  
or anyone on opposite sides of the 36° 30' parallel.

The issues stemmed from many years of discussion and careless thinking. Civil Rights were practically thrown out the door, and Civil War will soon be the horrible result.