CRESST REPORT 745

Christy Kim Boscardin
Barbara Jones
Claire Nishimura
Shannon Madsen
Jae-Eun Park

ASSESSMENT OF CONTENT UNDERSTANDING THROUGH SCIENCE EXPLANATION TASKS

NOVEMBER, 2008

National Center for Research on Evaluation, Standards, and Student Testing

Graduate School of Education & Information Studies
UCLA | University of California, Los Angeles

Assessment of Content Understanding through Science Explanation Tasks

CRESST Report 745

Christy Kim Boscardin, Barbara Jones, Claire Nishimura,
Shannon Madsen, and Jae-Eun Park
CRESST/University of California, Los Angeles

November 2008

National Center for Research on Evaluation,
Standards, and Student Testing (CRESST)
Center for the Study of Evaluation (CSE)
Graduate School of Education & Information Studies
University of California, Los Angeles
300 Charles E. Young Drive North
GSE&IS Bldg., Box 951522
Los Angeles, CA 90095-1522
(310) 206-1532

Copyright © 2008 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.

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

TABLE OF CONTENTS

ASSESSMENT OF CONTENT UNDERSTANDING THROUGH SCIENCE EXPLANATION TASKS

Christy Kim Boscardin, Barbara Jones, Claire Nishimura,
Shannon Madsen, and Jae-Eun Park
CRESST/University of California, Los Angeles

Abstract

Our recent review of content assessments revealed that language expectations and proficiencies are often implicitly embedded within the assessment criteria. Based on a review of performance assessments used in high school biology settings, we have found a recurring discrepancy between assessment scoring criteria and performance expectations. Without explicit scoring criteria to evaluate the language performance, it is difficult to determine how much of the overall performance quality can be attributed to language skills versus content knowledge. This is an especially important validity question for English Learners (ELs) under the current state assessment mandates. To date, studies of the validity and consequences of standards-based assessments for ELs have been limited. In the current study, we examined the various cognitive demands including language skills associated with successful performance on content assessments. Also, as part of the validity investigation, we developed and examined the relative sensitivity of performance-based assessment, which is constructed to be a more proximal measure of student understanding and sensitive to detecting instructional differences.

CHAPTER 1: INTRODUCTION

Background

Inclusion of English Learners (ELs) in high-stakes tests raises questions about the validity and reliability of these tests to inform decisions about accountability and instruction (August & Hakuta, 1997). Under the No Child Left Behind (2002) mandates, by the 2007–2008 academic year, all states are required to administer science content assessments to all students, including ELs. However, the Department of Education has not yet provided states and schools with adequate technical assistance and resources necessary to develop valid content-based assessments. To date, studies of the validity and consequences of standards-based assessments for ELs have been limited in number.

One concern stemming from the inclusion of ELs in high-stakes tests is the question of the validity of content assessments. Are content-based assessments actually measuring students' content knowledge and skills, or are these tests unintentionally assessing students' language proficiency? Without a clear determination of the validity of the standards-based assessments used in high-stakes testing for ELs, it will be difficult to evaluate the effect of the recent standards-based reform initiatives. One of the presumptive benefits of standards-based performance assessments is that they can more effectively measure complex thinking and problem solving. Measuring these skills may encourage schools to address standards related to thinking processes instead of focusing on the kinds of drill-and-kill basic skills that have often been associated with high-stakes multiple-choice tests. However, the validity of these types of measures for ELs is still in question. Another issue concerning the inclusion of ELs in high-stakes testing is the persistent achievement gap between ELs and non-ELs (Abedi, 2004; Aguirre-Muñoz et al., 2005). As states struggle to provide and achieve high standards for all students as well as demonstrate yearly progress towards meeting these goals, understanding the factors attributed to performance gaps will be paramount. In addition, our review of state content assessments revealed that language expectations and proficiencies are often implicitly embedded within the assessment criteria. Currently, 27 of the states require either an extensive or short written constructed responses as part of their statewide science assessment. Given the implicit language demands required by these assessments, without explicit scoring criteria to evaluate the language performance, it is difficult to determine how much of the overall performance quality can be attributed to language skills versus content knowledge.

In addition, previous studies have shown that the proximity or the degree of alignment of assessments to curriculum content also impacts the evaluation of instructional sensitivity

(Aguirre-Muñoz et al., 2005; Ruiz-Primo, Shavelson, Hamilton, & Klein, 2002). According to Popham (2006), instructional sensitivity of test refers to the degree to which students' performances on that test accurately reflect the quality of instruction that was provided to promote students' mastery of the skills assessed by the test. As part of the validity check, assessment must be able to differentiate the impact of various instructional qualities. This type of validity evidence will be paramount for the current accountability system that relies significantly on test scores to make judgments about instructional and educational quality.

For the current study, we propose to examine the various cognitive demands including language skills associated with successful performance on content assessments. Also, as part of the validity investigation, we will examine the relative sensitivity of performance-based assessment which is purported to be a more proximal measure of student understanding compared to standardized assessment such as the California Standards Test which represents a more distal measure on detecting instructional differences.

Content Assessment and ELs

There has been growing concern over the validity of content assessments with EL students, as previous research studies have shown that students' performance on content area assessments is confounded with language proficiency (Abedi & Leon, 1999; Bailey, 2000; Butler & Castellon-Wellington, 2000). Bailey suggests, based on a linguistic analysis of current standardized tests in content areas, that the language demand of a content assessment is a potential threat to the validity of the assessment and may not provide an accurate picture of EL students' content knowledge. As acknowledged in the *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999), the language demand of content assessments may introduce construct-irrelevant components into the testing process for EL students. Furthermore, it was concluded that these test results may not accurately reflect the competencies which are intended to be measured. Additionally, previous research suggests that most content assessments conducted in English inadvertently function as language proficiency tests rather than assessments of content knowledge (Abedi, Leon, & Mirocha, 2003).

To help minimize the effect of construct-irrelevant factors on the assessment of EL students, accommodations have been suggested and used in large-scale state assessments. However, sources of evidence for validity and effectiveness of these accommodations in reducing language factors in content assessments have been limited (Abedi, Courtney, & Leon, 2003). Also, certain types of accommodations are irrelevant and difficult to implement in the context of performance assessments. For example, allowing extra time and use of a dictionary may help reduce cognitive and linguistic loads in a standardized testing format (i.e., multiple-choice tests); however, for performance assessments that typically include reading and writing components, extra time and the availability of a glossary are part of general assessment guidelines and may not provide additional accommodations for ELs.

Additionally, a problem may arise when rating the quality of student work on performance assessments if the evaluation criteria do not distinguish the language demands of the task from the content demands, making it difficult to determine how much of each of these factors contribute to the overall score and impression of student performance. McCutchen (1986) posits that students' linguistic sophistication in writing (i.e., ability to demonstrate control over the linguistic resources needed to realize a particular genre) can compensate for a lack of content knowledge, but high levels of content knowledge do not necessarily compensate for a lack of writing proficiency. A high level of content knowledge certainly does not, in itself, work to the detriment of student writing. If, however, a writer does not understand how texts are constructed or how to use language in a text to lead a reader through an argument or an explanation of a process, then the value of the writer's content knowledge may be weakened because the writer cannot communicate this knowledge effectively to others. This notion is particularly relevant and potentially problematic for ELs when the assessment format requires responses in writing.

Our evaluation of commonly used performance assessments also reveals that language expectations are often implicitly embedded within the assessment criteria. Based on a review of performance assessments used in high school biology settings, we have found a recurring discrepancy between assessment scoring criteria and performance expectations. For example, in the Advanced Placement (AP) Biology exam scoring guidelines, points are awarded to student writing based on the inclusion of certain content information. The final score is the accumulation of these points. However, we found that the student prompts for the AP Biology writing exam ask students to *describe*, *discuss,* and *explain* the biology content, whereas the scoring criteria do not explicitly take into account students' varied ability to undertake these communicative writing processes which demand specific academic language competence. For example, to write a clear description of a biology concept, students need to include specific content vocabulary that is condensed into expanded noun phrases (a noun plus any modifying information, such as "*the vast number of <u>cells</u> needed for the embryo to take form*"). Thus, it is unclear whether the raters' scores clearly measure content understanding or a combination of content understanding confounded with the students' language proficiency for describing and explaining.

Current biology textbooks also frequently include writing components as part of the regular chapter assessments. For example, in the 2006 Prentice Hall *Biology* textbook, (Miller & Levine, 2006) students are asked to complete writing assessments called "Writing in Science" as part of the end-of-chapter assessments. This task entails writing a paragraph or group of paragraphs on target biology content. Like the AP Biology writing exam, Prentice Hall's writing assessment criteria explicitly refer only to the scoring of content knowledge. For example, in response to a prompt in which students are asked to write a paragraph that includes: (a) an explanation of a polymer, (b) a description of organic compounds, and (c) how these organic compounds are used in the human body, the evaluation criteria relate only to biology content (e.g., one of the criteria requires that students "explain that a polymer is a macromolecule made up of monomers joined together"). None of the evaluation criteria pertains specifically to the language features needed to successfully provide an explanation of the content. As with the AP Biology Exam, students are expected to communicate science concepts through specific academic language processes, though these processes are only implicitly evaluated as part of the assessment score. Without explicit scoring criteria to evaluate language, it is difficult to determine how much of the performance quality can be attributed to language skills versus content knowledge.

An additional measurement concern associated with typical writing assessments is the use of holistic rubrics to rate student writing and content knowledge together. Holistic rubrics usually include a variety of criteria (e.g., content, organization, spelling, grammar, etc.) and result in a single, overall score of the writing quality. An example of this is the holistic rubric used to score the Scholastic Aptitude Test (SAT) writing assessment, which rates student writing in consideration of five different points. Roughly, these five points are: (a) development of a point of view with clear evidence and use of critical thinking; (b) focus, smooth progression of ideas, and good organization; (c) skillful use of language with varied, accurate, and apt vocabulary; (d) meaningful sentence structure variety; and (e) correct grammar, usage, and mechanics. This type of rubric allows raters to rate the overall effect of the writing and its success in communicating ideas to the reader.

The rubrics used to score the SAT allow a rater to explicitly consider several aspects of student writing that contribute to its quality. In cases where a student does not receive a high score, however, a holistic rubric does not provide enough information to determine whether the student was lacking in academic language proficiency or in their understanding of the target concepts. Our rationale and conceptual framework for a new approach to assessing content understanding using an already widely accepted and utilized performance assessment format are discussed next.

New Assessment Framework

For this study, it was desirable to choose a context in which we could assess students' academic language proficiency as well as specialized content knowledge in the format of a performance assessment. To this end, we chose high school biology classrooms as the context for this study. High school biology classrooms are appropriate for this endeavor because they tend to rely on science texts as a resource for gaining and expressing science knowledge.

As indicated by the *National Science Education Standards* (National Research Council, 1996), students need opportunities to present their understanding and to use knowledge and the language of science to communicate scientific explanations and ideas. Grounded in the *National Science Education Standards*, we define acquisition of conceptual understanding in science to include not only the mastery of particular concepts and ideas but also the ability to effectively integrate concepts into the formulation of persuasive explanations for particular problems or phenomena. The National Center for Research on Evaluation, Standards, and Student Testing's (CRESST) model-based assessment uses standard architectures embedded in disciplinary content to assess core types of learning—basic knowledge, conceptual understanding, problem solving, communication, and teamwork. Two different templates have been developed and extensively validated as a way to observe science content understanding and communication, one using writing explanations given primary source materials, and the other using computer-based knowledge mapping to display comprehension (Baker, 1994; Chung, O'Neil, & Herl, 1999). As described below, CRESST's explanation architecture for assessing basic content knowledge and conceptual understanding provides a primary framework for the proposed assessment in the current study.

The explanation architecture seems particularly suitable for the current study context in that science is about the construction of theories that *explain* how the world operates, and scholars have noted that discourse, explanation, and argumentation are at the heart of science learning (Boulter & Gilbert, 1995; Duschl & Osborne, 2002; Erduran, Simon, & Osborne, 2004; Pontecorvo, 1987). The explanation architecture provides a format for examining whether students are able to integrate a complex structure of biological concepts, including related concepts, relationships between concepts, reasons for these relationships, and ways to explain and predict other natural phenomena. Within school-based science writing, explanation tasks are a dominant genre of student writing (Martin & Miller, 1988) and lend themselves easily to on-demand assessment conditions. Moreover, science explanation, as a type of formal technical writing, is important in the development of skills that logically organize scientific and technical knowledge into explanatory strings which utilize an appropriate science register.[1] When students are provided with increased opportunities to practice the genre of explanatory writing, they gain competence in assembling explanations that use a proper science register and can move toward functioning as more confident participants in the scientific and technical subject areas.

In the proposed assessment model, the use of an explanation task with reading prompts and a writing component provides us with an integrated approach to measuring students' ability to understand and communicate their science knowledge, using appropriate literacy skills to convey that knowledge. Under this model, learning science also becomes learning to use the specific linguistic features and structures that help communicate scientific principles, knowledge, and ideas (Fang, 2004).

Using this proposed assessment model, in this study we are interested in: (a) accurately measuring ELs' content knowledge in performance assessments, (b) gaining a better understanding of whether this measurement of content knowledge is confounded with measurements of language proficiency (i.e., the ability to write using the target science register), and (c) exploring how these measurements relate to impressions of overall essay quality. Based on these goals, we developed three separate rubrics to evaluate student responses. The three rubrics include: (1) the *holistic* or overall quality of student writing, (2) student understanding of the target science *content*, and (3) student ability to express scientific ideas in the target register using appropriate academic *language*. Each student essay, then, received three separate scores: Holistic, Content, and Language.

Rating student writing based on three separate rubrics allows us to determine whether, in general, strengths or weaknesses in student writing are more closely related to biology content understanding or academic language proficiency. Secondly, scoring in this manner allows us to discern whether raters' impressions of the overall quality of student writing is more closely aligned with students' content understanding or with their use of language. This will ultimately shed light on whether rating systems based on content, such as the AP Biology exam, or on a holistic system, such as the SAT writing exam, are valid measures for assessing EL content understanding and whether an explicit rating of language is needed in order to accurately measure written performance assessments.

Development of the Integrated Learning Assessment

The assessment we developed is called the Integrated Learning Assessment (ILA). For the purposes of research and validation, we developed four versions of the ILA, two in

[1] *Register* refers to "the configuration of lexical and grammatical resources which realizes a particular set of meanings" (Schleppegrell, 2004, pp. 45–46)

genetics ("Genetics 1" and "Genetics 2") and two in physiology ("Physiology 1" and "Physiology 2"). Each contains a reading passage, questions on students' reading processes, multiple-choice reading comprehension questions, and a writing prompt. The theoretical framework for the development of the ILA is described next.

As part of the development work, California state content standards for Biology/Life Sciences were reviewed extensively. The content standards are designed to define the knowledge, concepts, and skills that students are expected to acquire at each grade level or grade-level span. The concept of biology, which informs the standards, is described in the California Department of Education's (CDE) Science Curriculum Framework. The CDE conceptualization is consistent with generally accepted notions of biology. The field of biology is conceptualized as the study of life. The standards describing the field of biology focus on the common qualities of various life forms and the idea of the organism. In the CDE framework, the study of biology is organized into sections consistent with sub-fields of scientific study, such as ecology, physiology, and evolution.

Ontology

As part of the development process, after studying the biology standards, we created an ontology—a systematic arrangement and categorization of concepts in a field of discourse. Developing this type of organizational structure allowed us to uncover the relationships between different biology concepts (e.g., what concepts encompass the precursor knowledge set needed to understand a specific standard). This involved unpacking and elaborating the standards to create a hierarchy of conceptual information. This hierarchy of information was then used to: (a) create a framework for content understanding, (b) shape the design of the ILA, and (c) guide the development of the content rubric. Using the California Science Teachers Association's *Making Connections: A Guide to Implementing Science Standards* (Bertrand, DiRanna, & Janulaw, 1999) as a guide, we examined the standards in two specific science content areas—genetics and physiology. We determined what prior content knowledge would be necessary to understand the target standards, based on preceding units, content standards and sub-standards, and science standards from preceding grade levels (e.g., California Science Standards for Grade 7).

The two sections of Biology targeted for the ILA are genetics and physiology,[2] both of which are well represented in the California Science Test (CST) for biology. We chose these two topics for the ILA because we expected teachers to spend more time covering these content areas during classroom instruction, permitting us to relate opportunity-to-learn (OTL)

---

[2] see http://www.cde.ca.gov/be/st/ss/scbiology.asp

variables with student performance on science achievement tests. Overall, the review of content standards and the development of a biology ontology provided us with the context for determining the specific content targeted in the ILA.

Approach to Language Evaluation

In order to develop the various language-based assessment tasks in the ILA, reading and writing standards were examined, as well as concepts from systemic functional linguistics. The English Language Arts standards represent the primary guide for this development, whereas the systemic functional linguistic approach to language provides supplementary information to define distinctive features of academic language in science, which are different from those of informal spoken language. Using this framework, we were able to isolate and evaluate the specific linguistic features that are critical for proficiency in scientific literacy.[3]

In order for students to become scientifically literate, they must be able to read and comprehend scientific texts; this includes understanding the highly specialized academic language of these texts. They also need to be able to employ these same academic language resources to communicate what they have learned. Writing in science is realized through the appropriate use of a distinct set of lexicogrammatical resources (Fang, 2004). Using the functional linguistic approach to the analysis of academic language in science, we can systematically evaluate the effective use and control of the unique linguistic forms and structures, which are specific to communicating scientific knowledge.

The functional linguistic approach has also been applied in the past to analyzing students' writing performance, such as description (Schleppegrell, 1998, 2003), narrative (Christie, 1986), scientific essay (Christie, 1986; Schleppegrell, 2003), literary analysis, opinionated text (Christie, 1986, 2002), and response to literature (Aguirre-Muñoz et al., 2005). Functional linguistic analyses of student writing reported in previous studies reveal that students often lack understanding of expected language use in performing given academic tasks. For example, in a description task, students often invoke a non-academic interpersonal context (i.e., a situated, personal context) by employing the past progressive tense and first person references (Schleppegrell, 1998). These qualities indicate students' difficulties in producing the linguistic features, which are within the realm of academic writing.

---

[3] In keeping with the California Content Standards in Language Arts, an *explanation* may be alternately referred to here as an *essay*.

Description of ILA Development

Guided by these two approaches to language and content testing, our particular test specifications are closely aligned with the criterion-referenced measurement (CRM) paradigm. The purpose of utilizing this mode of test development was to have a measure that was independent of reference to the performance of other students.

The test specifications followed a precise format for all four versions of the assessment (Davidson & Lynch, 2002). Table 1 gives a brief description of this format. This format is discussed in more detail in the following sections (see Appendix A for the ILA test specifications).

Table 1

Test Specification Format

Specification	Description
General description	The general description is a detailed description of what is to be tested.
Prompt attribute	The prompt attribute is a detailed description of what is to be given to the test takers and what they will need to do. This includes the test format (multiple choice, essay, short answer).
Response attribute	The response attribute, though somewhat overlapping with the prompt attribute, is a description of what should occur when the test takers are given the prompt attribute.
Sample item	The sample item is an example of the specific format and content patterns for test items that are written from the test spec.

The first phase in developing the test specification for the ILA was to provide a detailed description of what was to be tested. It was determined, based on the review of the standards that the ILA should incorporate high-level cognitive skills such as analysis, elaboration, evaluation, and synthesis of ideas learned in biology class combined with the ideas presented in the ILA reading passages. The tasks in the ILA were aimed at eliciting students' use of higher-level cognitive skills when engaging in reading and responding to the text through writing.

A specialized general description (GD) was constructed for each of the four versions of the ILA. Each version of the ILA included unique biology content standards to be evaluated, though the reading and writing standards were the same for all versions. The standards on which the GDs were based can be viewed in Figures 1– 3 below.

Biology/Life Sciences Standards

ILA: Genetics 1

Standard 5: The genetic composition of cells can be altered by incorporation of exogenous DNA into the cells.

Sample basis for understanding this concept:

5a. Students know the general structures and functions of DNA, RNA, and protein.

5b. Students know how genetic engineering (biotechnology) is used to produce novel biomedical and agricultural products.

ILA: Genetics 2

Standard 4: Genes are a set of instructions encoded in the DNA sequence of each organism that specify the sequence of amino acids in proteins.

Sample basis for understanding this concept:

4a. Students know the general pathway by which ribosomes synthesize proteins.

4d. Students know specialization of cells in multicellular organisms is usually due to different patterns of gene expression.

ILA: Physiology 1

Standard 10: Organisms have a variety of mechanisms to combat disease.

Sample basis for understanding this concept:

10a. Students know the role of the skin in providing nonspecific defenses against infection.

10b. Students know the role of antibodies in the body's response to infection.

ILA: Physiology 2

Standard 9: As a result of the coordinated structures and functions of organ systems, the internal environment of the human body remains relatively stable (homeostatic) despite changes in the outside environment.

Sample basis for understanding this concept:

9b. Students know how the nervous system mediates communication between different parts of the body and the body's interactions with the environment.

9c. Students know how feedback loops in the nervous and endocrine systems regulate conditions in the body.

Figure 1. Biology/life standards targeted in the ILA

ILA: (All)

Reading Comprehension Standards:

2.0 Read and understand grade-level appropriate material. Analyze the organizational patterns, arguments, and positions advanced.

Structural Features of Informational Materials

2.5 Extend ideas presented in primary or secondary sources through original analysis, evaluation, and elaboration.

Figure 2. Reading comprehension standards targeted in ILA

ILA: (All) – Writing Standards

Writing Strategies:

1.0 Write coherent and focused essays that convey a well-defined perspective and tightly reasoned argument. The writing demonstrates students' awareness of the audience and purpose.

1.1 Establish a controlling impression or coherent thesis that conveys a clear and distinctive perspective on the subject and maintain a consistent tone and focus throughout the piece of writing.

1.2 Use precise language, action verbs, sensory details, appropriate modifiers, and the active rather than the passive voice.

Writing Applications:

2.0 Combine the rhetorical strategies of narration, exposition, persuasion, and description to produce texts of at least 1,500 words each. Student writing demonstrates a command of standard American English and the research, organizational, and drafting strategies.

2.3 Write expository compositions, including analytical essays and research reports

a. Marshal evidence in support of a thesis and related claims, including information on all relevant perspectives.

b. Convey information and ideas from primary and secondary sources accurately and coherently.

c. Anticipate and address readers' potential misunderstandings, biases, and expectations.

d. Use technical terms and notations accurately.

Written and Oral English Language Conventions:

1.0 Write and speak with a command of standard English conventions.

Grammar and Mechanics of Writing

1.1 Identify and correctly use clauses (e.g., main and subordinate), phrases (e.g., gerund, infinitive, and participial), and mechanics of punctuation (e.g., semicolons, colons, ellipses, hyphens).

1.2 Understand sentence construction (e.g., parallel structure, subordination, proper placement of modifiers) and proper English usage (e.g., consistency of verb tenses)

1.3 Demonstrate an understanding of proper English usage and control of grammar, paragraph and sentence structure, diction, and syntax.

Figure 3. Writing standards targeted in the ILA

Prompt Attribute: Text Selection

For the purpose of choosing the reading passages for the ILA, it was necessary to have a clear understanding of what linguistic resources are used to create scientific meaning and what level of reading comprehension proficiency is required at the high school level. To gain this understanding, and following a functional linguistic schema, we conducted a linguistic analysis of high school biology textbooks, Prentice Hall's *Biology* (Miller & Levine, 2006) and BSCS's *BSCS Biology: A Molecular Approach* (Greenberg, 2001). The results of the analysis were used as a basis for selecting and modifying the texts used in the ILA (see Appendix B for a more detailed analysis).

Overall, we found that high school science textbooks display high technicality and abstractness. This was evidenced by the frequent occurrences of technical vocabulary and abstract nouns. In addition, various instances of "grammatical metaphor" (Halliday, 1994) were identified in biology textbooks.[4] For example, experiential information (i.e., what is happening in the text) was frequently expressed in nominal groups through nominalization (e.g., forming the noun "invasion" from the verb "invade"). These nominal groups were further expanded through the addition of an embedded clause, an adjective, or a prepositional phrase, resulting in high lexical density. Relationships between experiential elements were marked through various connectors including conjunctions, but verbal groups often subsumed the marking of conjunctive relationships (e.g., "to be followed by" instead of "and then").

This comparative analysis helped us select the four final passages to include in the ILA. These passages were similar to the textbook passages in terms of the linguistic difficulty. The final four text passages were from Internet sites[5] listed as supplemental resources for students in state-adopted biology textbooks (Miller & Levine, 2006; Campbell & Reece, 2005).

After the text passage selection, we conducted an external review of the text passages. This step involved consulting with genetic scientists at the University of California, Los Angeles, (UCLA) for content accuracy, as well as communicating with the authors of the text passages to obtain permission for use and to confirm that the content of the passage could still be considered current and accurate. The text passages were also reviewed and holistically rated by current high school biology teachers for us to gain a sense of the level of difficulty the texts and prompts might pose for their students (see Appendix C for the evaluation guide that teachers used to rate the texts and items).

---

[4] A *grammatical metaphor* is a process whereby meaning is constructed in the form of incongruent (i.e., metaphorical) grammar. Incongruency is characteristic of written discourse in relatively formal registers.

[5] http://citnews.unl.edu/nutrition/html/lesson.shtml?lesson_id=991751218)
http://sciencenow.sciencemag.org/cgi/content/full/2006/113/2
http://www.intelihealth.com/IH/ihtIH/WSIHW000/7945/8218/174513.html?d=dmt
http://web.sfn.org/content/Publications/BrainBriefings/mind.body.html

Item Generation

The generation of items for the ILA (i.e., writing prompts and reading comprehension questions) followed the text selection and was a multi-step process. We included a reading comprehension section in the ILA as a way to determine whether the quality of students' writing responses was influenced by their reading comprehension levels. The multiple-choice questions were developed with three categories in mind: vocabulary verification, sentence paraphrasing, and global paraphrasing. After the generation of many reading comprehension questions, three were selected to be included in each version of the ILA.

The development of the ILA also involved creating several candidate essay prompts based on the content of the text and science curriculum, with the requirement that they elicit higher-order thinking skills. Our desire was to create prompts related to each reading passage that could be adequately answered by utilizing science content learned that year in Biology class together with information directly gathered from the text passage. Once several potential prompts were developed for the ILA, they were reviewed and evaluated, and the final four prompts were selected. To verify the appropriateness of the text passages, reading comprehension questions, and writing prompts, we conducted a focus group with a small group of 9th-grade students (n = 5). The results from this process indicated that students of varying competency levels would most likely be able to respond to the various sections of the ILA. In the reading process section, all participating students in this screening responded that the text passages were between "easy" and "not too difficult to understand." The description of the final four versions of the ILA is provided in the Methods section and also can be found in Appendix D.

Research Questions

In this study, we propose to examine the validity issues associated with performance assessment when measuring content understanding for ELs and to examine the feasibility of differentiating different cognitive demands of content assessments. The study will be guided by two principal research questions:

1. Are we able to differentiate the various cognitive demands associated with successful completion of science explanation tasks using the newly developed assessment model? Specifically, what proportion of content versus language skills can be attributed to the overall evaluation of student performance?
2. What is the validity of ILA as an outcome measure? Specifically, is ILA more sensitive to detecting instructional differences compared to the other standards-based assessment?

In order to address these two research questions, we developed two new instruments specifically for the study: ILA and teacher assignment rubrics. As presented in the previous section, ILA was developed to provide a mechanism to systematically study various cognitive demands. Also, the teacher assignment rubrics were developed as an indicator of instructional quality. More detailed information regarding the teacher assignment rubrics are provided in the Methods section. The data collected from this study will not only address the two research questions but also serve as pilot data for instrument validation. One of the caveats to this approach is the limitations of the data for interpretation and generalizability of the findings.

CHAPTER 2: METHODS

In this chapter, we provide a description of the research design, instrumentation, and analyses used to address the research questions. We present the methods used for analyzing student assessment, teacher assignment, and classroom observation data. Also, we briefly describe how the different data sources were utilized in the statistical analyses to address the research questions.

Participants

Teacher

Twenty-six teachers participated in this study from 14 high schools in the greater Los Angeles metropolitan area. There were 12 public schools from 6 districts, and there were 2 private schools. Of the 26 participating teachers, 19 were women and 7 were men. Their length of teaching experience ranged from 1 to 34 years, with an average of 9.39 years.

To measure student OTL in relationship to student performance on the ILA, teachers were asked to submit two typical in-class biology assignments (one from ***Genetics*** and one from ***Physiology***) that involve some reading, with 6 samples of student-produced work for each assignment. Teachers were also asked to fill out a cover sheet for each assignment. The cover sheet was used to gather detailed description of the biology assignments that students were asked to complete and the evaluation criteria used to judge student work. We received biology assignments from 22 teachers. Four of the teachers administered the student assessment only and did not submit their in-class biology assignments.

Student

A total of 1,167 students were administered the exam by 26 participant teachers. Teachers were randomly selected to administer either genetics or physiology prompts. The breakdown by prompt is as follows: 299 (25.6%) students were administered the Genetics 1 prompt (11 classes), 277 (23.7%) were administered the Genetics 2 prompt (12 classes), 313 (26.8%) took Physiology 1 (13 classes), and 278 (23.8%) took Physiology 2 (12 classes).

Of the 1,167 students, 503 (43.1%) were female, 433 (37.1%) were male, and 231 (19.8%) did not report gender. 68 (6%) students were classified as English Learners and the EL status was unknown for 234 (20.1%) students. There were 94 (8.1%) 8th-grade students, 562 (48.2%) 9th graders, 235 (20.1%) 10th graders, 42 (3.6%) 11th graders, and 234 (20.1%) students whose grade was unknown.

For the statistical analysis, students enrolled in private high schools were not included in the final analysis because these students did not participate in California STAR testing program. Also, students enrolled in the classroom of teachers who did not submit their in-class biology assignments were also excluded in the final statistical analysis because these students did not have matching teacher information. As a result, a total of 755 students were included in the final analysis.

Procedure and Instruments

Description of the ILA

The ILA was designed to evaluate students' reading processes, reading comprehension, language proficiency in the written explanation genre, and science content understanding.

The CRESST research team developed four versions of the ILA, two in genetics and two in physiology, each consisting of a text passage followed by sections for short-answer reading process questions, multiple-choice reading comprehension items, and an essay prompt. The topics of the texts were in the areas of genetic modification techniques, genetic identification, immunological responses, and the mind-body link. The open-ended reading process questions were part of the CERA[6] assessment previously developed by WestEd's Strategic Literacy Initiative and measure the quality of reading strategies students employ to understand the text. The multiple-choice section relates to the content of the text and was constructed to measure specific attributes of reading comprehension, such as vocabulary knowledge and overall text understanding. Students were allowed to refer back to the text while working on this section. Each essay prompt was designed as a two- or three-part question. The essay prompts required students to draw on their prior content learning (English Language Arts and biology), recall information from the text they had read, and utilize higher-order thinking skills to complete the essay. Following each essay question, students were reminded to include the most important aspects of the content in relation to the questions and to pay attention to organization when completing the essay.

Scoring the ILA

Because one of the principal aims of this study is to examine both language proficiency—related constructs and biology content—related constructs; separate rubrics were developed to evaluate biology content knowledge, academic language proficiency in the student explanations, and overall writing quality. The three resulting rubrics were rated on a 4-point scale and titled Content, Language, and Holistic. Through the characterization of

[6] CERA refers to the *Curriculum Embedded Reading Assessment*

their respective score points, these rubrics describe various proficiencies of learning. Specifically, these rubrics measure levels of overall proficiency in writing in the explanation genre (Holistic), biology content knowledge (Content), and academic language proficiency (Language). Each score point within a given rubric provides a portrait of a student's biology explanation as it may appear at a given proficiency level (See Appendix E for the ILA rubrics).

Holistic rubric. For the Holistic rubric, we modified the Language Arts Performance Assignment holistic scoring rubric that was developed and validated in previous CRESST work (Aguirre-Muñoz et al., 2005) to be able to use in a biology content setting. The Holistic rubric measures a variety of criteria related to global aspects of the science explanation genre. The main components of the modified Holistic rubric were: understanding of key biology concepts, effective communication, and overall organization. A student response that received a high score on this rubric had to satisfy most or all of these criteria, elaborated in the rubric as follows: "the response demonstrates in-depth understanding of the relevant and important ideas, effectively communicates an explanation of the relevant scientific process, has strong logical organization that enhances the central ideas of the essay, and uses correct paragraph structure." Additionally, it was in the Holistic rubric that raters were to take into account whether student responses included mechanical errors that impeded communication, contained irrelevant information, or did not address the question being asked in the prompt. Because the Holistic rubric is intended to measure the overall quality of student responses, the criteria intentionally overlap somewhat with those of the other two rubrics used to score student responses. The Content and Language rubrics, then, provide more detailed accounts of specific student strengths and weaknesses, shedding light on which strengths and weaknesses have the greatest impact on perceptions of overall writing quality. Table 2 illustrates the constructs measured in the Holistic rubric, how they are realized in the explanation genre generally, and how they are realized in science explanation writing specifically (which students need to demonstrate in order to receive a high score point). Although the biology content does not change depending on the genre, this framework enables us to discern how the overall organization of the writing becomes more specialized in science explanation writing.

Table 2

Description of Holistic Rubric Constructs

Holistic constructs	General qualities of explanation genre	Specific qualities of science explanation genre
Genre elements and structure: Appropriateness of text organization to realize genre, including structures that create logical and clear progression of ideas.	**General language features** Paragraph- and sentence-level structuring of ideas provides: Sentences and paragraphs ordered/grouped by function, (e.g., to provide introduction, background, statement, argument, evidence, etc.)	**Specific language features** Paragraphing and sentence structuring provides: Presentation of a thesis with supportive evidence. Interpretations and generalizations to account for how and why things are as they are. Logical organization that presents causes and makes judgments. Emphasis on generalization, classification, and categorization of support documents.
Biology content: Demonstration of understanding of important ideas elicited in prompt.	**General biology content** <u>Example: Genetics 1</u> Students should know the general structures and functions of DNA and protein: A gene is made up of DNA. DNA is deoxyribonucleic acid. DNA codes for amino acids that make up proteins.	**Specific biology content** <u>Example: Genetics 1</u> DNA is the genetic instruction specifying the biological development of all life forms. It is a long strand of nucleotides and codes the sequence of amino acids in proteins. Proteins establish an organism's phenotype and genotype.

Content rubric. For the Content rubric, we included criteria that were deemed important as overarching measures of students' understanding of biology content. These criteria were formed, in part, by using the previously developed and validated CRESST rubrics as a guide. Our goal, which was to measure students' knowledge of concepts, ability to connect principles and concepts, and ability to extend prior knowledge of concepts beyond the limited contexts in which they were acquired, led us to develop a list of four initial key points upon which to base our rubric. The key points were: understanding of the target biology content, clarity of explanation, use of supportive evidence, and inclusion of prior knowledge. We referred to our biology ontology to ensure that these criteria were aligned with the target biology standards.

A student response receiving a high score on the Content rubric had to satisfy most or all of the criteria. These concepts were elaborated in the rubric's 4-point description as follows: the response demonstrates well-developed understanding of the target biology content and the content is clear, focused, thoroughly explained, and elaborated with strong supportive evidence. This dimension also encompassed whether or not a student demonstrated knowledge of the relevant content beyond explicit information given in the text passage (i.e., whether or not a student incorporated prior knowledge). The specific content raters were to look for in student responses was elaborated in the supplemental documents for the Content rubric, which were generated from the biology ontology.

Language rubric. In developing the Language rubric, we modified the Language dimensions that were previously developed and validated in earlier CRESST work (Aguirre-Muñoz et al., 2005) in order to align them with a biology-content setting and the explanation genre. Key points were used to evaluate students' academic language proficiency on the ILA based on the linguistic dimensions identified by the systemic functional linguistic approach utilized to operationalize academic language as noted earlier. The Language rubric specifically focuses on assessing students' linguistic command of grammatical structures that are directly related to the explanation genre in general and to the science explanation genre in particular and that are aligned with the California Content Standards in writing (See Appendix F for tables that elaborate the standards mapping). The language qualities of science writing that are of interest to us are abstraction, informational density, and technicality.

Based on previous CRESST work (Aguirre-Muñoz et al., 2005), we knew that most students in the early years of high school do not have the academic language proficiency to produce high-quality, academic explanations in science. For this reason, the Language rubric was structured to be sensitive to measuring a range of academic language proficiency levels in science writing. We related the ideas of abstraction, informational density, and technicality to three systemic functional linguistic concepts. *Mode* (the manner in which ideas are communicated) refers students' ability to create appropriate text cohesion in their writing, *Field* (the linguistic elements used to communicate them) refers to students' ability to use varied and precise word choice, and *Tenor* (the tone of that communication) refers to students' ability to establish a formal, impersonal tone in their writing.

In the Language rubric, as shown in Table 3, we operationalized the idea of field in science explanation writing in terms of varied and precise word choice, including technical terminology. We also looked for expanded word groups, especially expanded noun phrases, to describe the biology concepts.

Varied and precise word choice (Field). The Field of discourse is associated with presentation of ideas, typically involving "content" words such as nominal groups, verbal groups, and adverbial expressions. In science writing in particular, the dimension of field is characterized by *informational density*, whereby clauses carry a high percentage of content-specific words. These tend to be nouns, verbs, adjectives, and adverbs. Content words are usually clustered into phrases (e.g., expanded noun phrases, which can be used to condense information). This high use of content words and, at times, technical vocabulary leads to another characteristic of science writing, namely, *technicality*. This is realized through the use of noun phrases and verbs that show relationships between them (Fang, 2004). Table 3 below provides additional information about the elements of this dimension.

Table 3

Description of Language Rubric Dimension: Field

Construct: varied and precise word choice	Operationalized in the rubric as:	Specific language features used to realize field	Specific function of language features in science explanations
Information density and technicality	**Word group quality**: Variety and expanded-ness of word groups **Lexical quality** defined as significant and appropriate use of technical terminology	**Noun groups can consist of** main noun, adjectives, embedded clauses, prepositional phrases. **Verb groups can consist of** verbs, adverbs, prepositional phrases. **Adverbial groups include** adverbs, subordinate and participle clauses, prepositional phrases. **Word Choice** specific to the science domain	**Noun groups** are often events or happenings instead of personal noun groups. They also name points to be made (e.g., *There are three reasons that…*) **Verb groups** include frequent action and having/being verbs **Adverbial groups** rank and condense information through use of subordinate clauses

Formal and impersonal tone (Tenor). In science writing, Tenor reflects a convention of formal, written discourse. That is, personal opinions and stances should be presented in an authoritative and impersonal fashion. This requires the use of interpersonal resources including the declarative mood, modal verbs, and lexical choices that carry an implicit evaluative meaning rather than choices that resort to an emotional appeal (e.g., rhetorical questions) or explicit evaluative meaning (e.g., "I think that" and "I believe that"). In the Language rubric, as shown in Table 4 below, Tenor is operationalized by considering whether the text has a formal tone and portrays personal opinion implicitly. An author establishes a formal tone by using the linguistic resources of third person and passive voice and by avoiding speech markers ("well," "you know," "like," etc.).

Another consideration for Tenor is the speaker or writer's display of stance (i.e., judgment or interpretation) in the text. The premise is that the speaker or writer expresses his or her personal stance in consideration of the listener or reader. Thus, the display of stance involves various linguistic resources that create the interpersonal meaning. Such interpersonal choices include mood, modality, intonation cues (in spoken discourse), and lexical elements that carry an evaluative and attitudinal meaning. Table 4 provides additional information about the elements of this dimension.

Table 4

Description of Language Rubric Dimension: Tenor

Construct: Formal and impersonal tone	Operationalized in the rubric as:	Specific language features used to realize Tenor	Specific function of language features in science explanations
Tenor Authoritative quality	**Tone of text** Defined as formal and impersonal	Passive voice Third person Few uses of speech markers Few addresses to oneself or audience	Used to create an impersonal stance.
		Modal verbs and adverbs (e.g., *can*, *will*, *possibly*, *perhaps*, etc.) "It" constructions	Used to present claims as possibilities.
		Precise word choice of nouns, adjectives, verbs, and adverbs	Used to convey evaluation, (e.g., *responsible)*.

Text cohesion (Mode). The Mode of discourse refers to the way that language is structured in the social context in which it is used. The structure of a text reflects both linguistic and non-linguistic aspects of the social context, such as the availability of feedback between speaker and listener or between writer and reader. Linguistic resources that construe the textual meaning include cohesive devices such as conjunctions and connectors, clause-combining strategies, and thematic organization. In the rubric, we characterized this dimension as *text structure* in order to reflect the elements of grammar that realize the type and organization of text that serves a specific purpose. When rating the Language dimension, raters considered whether students used a variety of sentence structures, including marked themes (information in front of the subject used to link it to the previous clause). Table 5 provides additional information about the elements of this dimension.

Table 5

Description of Language Rubric Construct: Mode

Construct: Text Cohesion	Operationalized in the rubric as:	Specific language features used to realize mode in science explanations	Specific function of language features in science explanations
Mode With qualities of: Abstraction and Information density.	**Text cohesion** The flow between clauses and sentences.	**Text connectors** (conjunctions, adverbials, verbs). **Marked themes** (information in front of subject). **Thematic progress** Subject of one sentence connected to the predicate of a previous sentence, (e.g., nominalization).	Text connectors and marked themes link one part of the text to the next with cohesive ties, causal conjunctions and markers of contrast, classification, and logical sequence. They also include grammatical shifts for moving from general to specific and back again. **Thematic progress** Information ranked and condensed through use of clause organization and nominalization.

In order to receive a high score on the Language dimension, a student's explanation had to meet most or all of the following criteria: demonstration of very good text cohesion through regular use of sentence structure variety (specifically, through use of marked themes and nominalizations), consistent use of precise and varied word choice (specifically, through use of expanded noun phrases), and use of an impersonal and authoritative tone with few or no speech markers. Length of the student's paper was taken into consideration only to the extent that the writing needed to be long enough to provide evidence of academic language proficiency.

The Scoring Session

A total of 1,167 student responses were evaluated during the 8-day scoring session. To minimize rater bias, all identifying information (student names, teacher names, and school names) was removed from the student papers. Responses were randomly distributed and divided into packets containing 20 responses each.

ILA Scores. All raters underwent intensive training during Days 1 and 2 to introduce and practice scoring procedures, address questions, and ensure that the scoring rubrics were clear. All student responses were scored by two different raters to achieve greater consistency. For reliability purposes, only scores within one point difference between the two raters were included in the data analysis. In the total sample, only about 8% of the cases had score disagreement of more than one point. The final scores for student responses represent the arithmetic mean of the two raters' scores. Although there are several methods for resolving rating disagreements, we opted to use the parity model by averaging the two scores.

Teacher Assignments

In order to determine the sensitivity of the ILA in detecting instructional differences, we measured the quality of teacher practice and students' opportunity to learn by collecting classroom assignments. The examination of classroom assignments has been developed and validated in previous studies as a reliable indicator of classroom practice (Aschbacher, 1999; Matsumura, 2000). Using classroom assignment ratings to assess the quality of classroom practice, provides an efficient and economically viable alternative to classroom observations. Assignment ratings also enable us to determine the types of instructional practices that are related to ILA student performance. The following section provides the theoretical framework informing our approach to evaluating literacy and science instructional quality in the classroom assignment ratings.

The CRESST criteria are based on research which indicates that effective teachers tend to include the following in their practice: (a) they maintain high standards for student achievement, (b) they hold clear goals for student learning, and (c) they align their assessment criteria with standards and goals for students (Aschbacher, 1999; Matsumura, 2000). We modified and adapted a previous rating rubric to specifically target science and literacy instruction. Our measures of literacy instruction are also informed by the Reading Apprenticeship® (RA) Framework, which is a model for literacy instruction in the content areas (Schoenbach, Greenleaf, Cziko, & Hurwitz, 1999). To inform our research question regarding instructional factors associated with higher performance on science explanation tasks, we decided to rate assignment quality on two separate rubrics, one to measure the quality of teacher practice in relation to science literacy and the other to rate the quality of teacher practice in relation to science content. For each rubric, we used a 4-point scale 1 (*poor*) to 4 (*excellent*) to rate the seven dimensions of quality for each assignment. A total of 14 dimensions (7 for content, 7 for literacy) were rated to evaluate the quality of teacher assignments, and these ratings serve as an indicator of instructional practice.

Based on the RA framework, we included criteria in our literacy rubric to measure teachers' integration of explicit literacy instruction, opportunities for engagement with text, and metacognitive processing into their teaching. These are aspects of science learning that are considered in this model to be necessary for comprehending advanced science concepts and developing science-related thinking skills.

Because the RA Framework is also centered on the notion that high quality classroom practice includes support for students in their efforts to meet the cognitive and metacognitive demands of their assignments (e.g., through use of appropriate literacy routines, material resources, and peer interaction), we decided to include a dimension in each rubric (literacy and content) to measure *support* for the assignment challenge. Inclusion of this dimension is based on the RA framework that characterizes effective teachers as appropriately supporting the challenge of the assignments in a manner that enables students to perform successfully while continuing to use critical thinking and metacognitive skills.

Overall, the two rubrics for literacy and content are very similar in most of the seven dimensions, except for the challenge dimension. In the literacy rubric, the challenge dimension focuses on the challenges presented by text engagement (including the use of reading comprehension strategies) and metacognition. In contrast, the challenge dimension in the content rubric focuses on levels of cognitive demand in relation to biology content (i.e., higher order thinking skills). A detailed description of both literacy and content rubrics is provided next.

Teacher Assignment Ratings: Science Content Dimensions

The following sections elaborate the qualities of the seven dimensions of the Content rubric, providing an overall description, explanation of the various score points, and samples of scored assignment descriptions for each content dimension.

Quality of the Goals for Student Learning. The purpose of this dimension was to describe how clearly a teacher articulated the specific scientific skills and biology concepts students were to utilize and gain from completing the biology assignment as well as to describe the degree to which an assignment could be considered to be focused on student learning in biology.

Rubric. This dimension was scored on three criteria: Purpose, Clarity, and Elaboration. Purpose considers whether the teacher's science content goals were specific and focused on student biology learning. Clarity considers whether the science content goals were clear and explicit in terms of what students were to learn as a result of completing the biology assignment. Elaboration considers the amount of detail in the description of the science content goals.

For example, in one biology assignment given a high score on this dimension, the teacher wrote that his or her “[s]tudents will know [that] random chromosome segregation explains probability of alleles becoming gametes, that half of all individuals’ DNA comes from each parent, determining sex-chromosomes X and Y, and how to predict possible phenotypes from genotypes.” For this particular assignment, the teacher’s goals for the students “to know” specific biology concepts and phenomena were stated. The biology concepts to be gained from completing the assignment are taken directly from the California Biology/Life Sciences standards for the 10th grade. Overall, the assignment had goals that were very clear, detailed, and specific as to the scientific skills and biology concepts students were to learn from completing the assignment. In contrast, the stated content learning goals for an assignment that received a low score on this dimension were that “[s]tudents should identify the three questions posed by the gene researcher,” and that “students should find examples of the use of genetic engineering in the movie (Gattaca).” This assignment was given a low score because it was not clear what the students were to do or whether these goals were intended to promote student learning. Instead, the assignment called for the students to recall ethical questions posed by a gene researcher on contraindications of genetic engineering in a society. The link between the goals for the assignment was not clear or purposeful.

Cognitive Challenge. The purpose of this dimension was to describe the degree to which students are required to apply complex science skills when engaging with scientific biology concepts. With regard to the opportunities to engage in complex science skills, this dimension considers the level of critical thinking required of the students in order to complete the assignment (e.g., critical thinking, problem solving, analyzing, and synthesizing information). Specifically, the dimension considers the opportunity students had to construct or transform knowledge as opposed to simply recalling, describing, or identifying basic information.

We considered the following three rating criteria when developing the Cognitive Challenge dimension of the content rubric: (a) the cognitive challenge had to be concerned with thinking about science content, (b) the cognitive challenge could be rated on the complexity of that content, and (c) the cognitive challenge rated the appropriateness of the level of thinking elicited by the assignment. The specific distinction between cognitive strategies and problem-solving strategies considered in other learning theories were combined and became embedded within our own conceptualization of this rubric dimension. For the content Cognitive Challenge, we considered the complexity of the content of the assignment. We wanted to establish that, although introductory science information is necessary for building the foundation of meaningful science knowledge and understanding, it was important for us to consider the depth of the science content learning taking place as marked by the complexity of the science content.

Rubric. This dimension was scored on the degree of higher-order thinking skills required of the students when engaging with biological concepts. This dimension also considers the depth of content understanding targeted through the Cognitive Challenge of the assignment.

An example of an assignment that received a high score on Cognitive Challenge is one in which students were required to analyze, synthesize, and interpret experiment results from a lab activity they designed, the process of which involved significant use of higher-order thinking skills. An example of an assignment that received a low score on this dimension, on the other hand, required students to read a chapter and answer three information recall questions about how communication between various parts of the body occurs. The students in the latter example were not required to utilize any higher-order thinking skills, such as analysis or synthesis, when engaged in this assignment. Students recalled basic science information in order to satisfy the requirements of the assignment.

Support for Cognitive Challenge. The purpose of this dimension was to describe how appropriately a teacher provided support for the cognitive challenge goals targeted by the assignment. Specifically, this dimension considers the degree of support for the thinking skills (e.g., knowledge, comprehension, application, analysis, synthesis, and/or evaluation) and processes that were provided by the teacher for successful completion of the assignment.

Rubric. Each score point was based on the degree of support provided to students through previous or current teaching approaches such as teaching thinking processes, structuring the cognitive activity into an appropriate number of explicit steps, enabling students to draw on peer knowledge, and making resources available to aid in meeting the cognitive challenge of the assignment.

An assignment that received a high score in this dimension required students to understand the process of gene splicing well enough to do it themselves with scissors, tape, and paper, each representing the different parts of the process. The teacher provided such support as direct instruction, visual aids (e.g., a flow chart), group work, and assistance during the activity. The support included a high degree of scaffolding in which the teacher "walked through the steps of gene splicing." This assignment received a high score because it provided support that was appropriate and focused on the cognitive task students were to carry out. In contrast, a class discussion held in support for a debate assignment received a low score because it was inappropriately scaffolded by the teacher. Much of the discussion was dominated by the teacher through detailed questions and answers being provided to the students by the teacher. Students had little opportunity to explore the content through their own discussion of false responses, beliefs about the content, or questions.

Quality of the Evaluation Criteria. The purpose of this dimension was to assess the quality of the content evaluation criteria for the biology assignment in terms of their specificity (i.e., how clear, explicit, and elaborated were the evaluation criteria) and their potential for helping students improve their performance and their understanding of biology concepts.

Rubric. Each score point was based on three criteria: Specificity, Information for Students, and Fidelity. Specificity considers whether teachers' content evaluation criteria were clear, explicit, and elaborated. Information for Students considers whether the criteria are detailed and communicated to students to improve their content learning. Fidelity of evaluation criteria considers whether the criteria were reflected in the graded selection of student work samples.

An assignment that received a high score on this dimension required students to draw a family tree and trace phenotypes. This assignment was awarded a high score for having stated science evaluation criteria that were clear and explicit. Students were to "[i]nclude at least 3 generations and number them, trace the phenotype in your pedigree, identify the genotypes for each individual, and classify and title the trait as modes of inheritance: autosomal dominant or recessive, or sex-linked dominant or recessive." Each section had a "points earned over points possible" column where students would be able to see where they had earned or lost credit. Additionally, the rubrics were returned to the students with their work, and it was apparent that the scoring rubric used was reflected well in the choice of work samples. In contrast, an assignment that did not score well on this dimension included the answer key to the handouts completed by the students. This form of evaluation criteria illustrates which answers students received credit for but does not contain enough detail to help them improve their understanding of the content. Also, there is no evidence to suggest that the students had the opportunity to view the grading key. Another assignment that received a low score on this dimension had stated evaluation criteria that did not address any science content.

Alignment of Learning Goals and Task. This dimension was focused on the degree to which a teacher's stated learning goals in biology are reflected in the design of the assignment tasks students are asked to complete. Specifically, this dimension attempts to capture how well the assignment appears to promote the achievement of the teacher's goals for learning in biology.

Rubric. Each score point was rated on the single criterion of alignment between goals and task. An assignment with a high score in this dimension was considered to have excellent alignment, with the task fully supporting the goals, neither calling for something not included in the other. On the other hand, an assignment receiving a low score in this dimension had very little alignment between goals and task, which resulted in the task providing little support for the achievement of the assignment goals.

For example, in one assignment the goals specified that students were to learn the following concepts: "identify how DNA technology can be used to improve lives, understand what restriction enzymes, recombinant DNA, and plasmids are; and correctly select restriction enzymes to create recombinant DNA." The assignment task required students to "read a scenario describing how hormone therapy might be used to treat certain conditions. The students then walked through the steps of how recombinant DNA technology can be employed to create the hormones. Students did a hands-on activity to mimic the splicing of genes, choosing the appropriate restriction enzymes and creating recombinant DNA." This assignment received a high score on this dimension because the assignment task fully supports the instructional goals for the science content. A hypothetical example of an assignment that would receive a low score on this dimension would be one with clearly stated learning goals for the students to understand the process of cell division and to be able to distinguish the roles of DNA and RNA during different stages of cell division. However, the task may only require the students to read a text section on cell division and to take notes in their notebooks. In this hypothetical assignment, there is only very general alignment between the goals and the task. It would be difficult to confirm that all the stated goals had been addressed by the assignment task.

Alignment of Learning Goals and Evaluation Criteria. This dimension was intended to describe the degree to which a teacher's evaluation criteria supported the learning goals in biology, that is, the degree to which a teacher assessed students on the science skills and biology concepts they were intended to learn through the completion of the assignment.

Rubric. Each score point was based on the single criterion of alignment between the goals and the evaluation criteria. An assignment with a high score in this dimension was considered to have excellent alignment, with the evaluation criteria fully supporting the goals, neither calling for something not included in the other. An assignment receiving a low score in this dimension had very little alignment between goals and evaluation criteria, which resulted in the evaluation criteria providing little support for the achievement of the assignment goals.

An example of an assignment that received a high score on this dimension had the following student learning goal: "To understand the cell organelles in terms of their interrelationships and specific roles." This assignment's evaluation criteria were clearly and explicitly aligned with the learning goals. The assessment criteria target, in detail, aspects of the learning goals (i.e., roles and relationships of the cell organelle, student understanding gauged by their ability to transform knowledge into different forms). Alternatively, an assignment that did not score well on this dimension did not include any evaluation criteria for the assignment task.

Overall Quality of Assignment. This dimension was intended to provide a holistic rating of the quality of the biology assignment based on its level of cognitive challenge, the support of student biology content learning, the quality of the goals for student learning, the quality of the grading evaluation criteria, the alignment of the learning goals with the assignment task, and the alignment of the learning goals with the grading evaluation criteria.

Teacher Assignment Ratings: Literacy Rubric Dimensions

The following sections elaborate the qualities of the seven dimensions of the Literacy rubric, providing an overall description, explanation of the various score points, and samples of scored assignment descriptions for each literacy dimension.

Quality of the Goals for Student Learning. The purpose of this dimension was to describe how clearly a teacher articulates the specific literacy engagement, skills, and thinking processes students were to utilize and/or learn for better content understanding as a result of completing the biology assignment. It also had to allow for evaluation of whether or not students had achieved these goals.

Rubric. This dimension was scored on three criteria: Purpose, Clarity, and Elaboration. Purpose considers whether the literacy goals are specific and purposefully focused on students' engagement with text, literacy related thinking processes, and use of literacy skills for learning. Clarity considers whether the literacy goals are clear and specific in terms of what students are to learn as a result of completing the biology assignment. Elaboration considers whether the literacy goals are described in detail.

In one biology assignment given a high score on this dimension, the teacher wrote that her "[l]earning goals include being able to do research using the internet, evaluating internet resources, writing a persuasive essay, and learning how to appreciate current events and making connection [*sic*] with the text book." In contrast, the stated literacy goals for an assignment that received a low score were: "follow instructions; planning and organizational skills [*sic*]; propose a hypothesis and identify variables." This assignment was given a low score for quality of literacy goals for student learning because it was not clear what students were supposed to learn from completing the task or what type of literacy engagement was intended to take place during the course of the assignment.

Literacy Challenge. The purpose of this dimension was to describe the degree to which this assignment task provides students with the opportunity to engage meaningfully with biology text and utilize metacognitive processes to reflect on and assess their understanding of the concepts within the text. With regard to engagement with text, this dimension also considered the richness of the text (written and visual information) and the appropriateness of the reading level for students.

Rubric. Each score point was based on three criteria: Complexity of Text, Engagement, and Metacognitive Challenge. To receive a 4, an assignment would need to include all three criteria. To receive a 3, the assignment had to, at a minimum, demonstrate some level of metacognitive challenge in relation to an appropriate text. Complexity of Text is considered in relation to a text's appropriateness for students' reading comprehension levels and its written and visual richness to allow for meaningful engagement. Engagement with science texts is considered in relation to the quality of the opportunity students receive to engage with texts. Metacognitive Challenge considers the degree to which students were required to reflect on their approach to the science reading task, their own science reasoning processes, and their understanding of science concepts derived from science texts (e.g., by identifying confusions or misconceptions through conversations with peers and then rereading the text).

An assignment receiving a high score on this assignment would, for example, require students to fill out a reading log during independent reading of a science text passage where they had to record their interpretations, questions, and connections to the text. Following this, the assignment may require them to participate with their peers in metacognitive conversations about their understanding of the text. This assignment would have received a high Literacy Challenge score because it provides students with opportunities to increase their understanding of the text through reflecting on how they become good readers of science. An example of an assignment that received a low score on this dimension asked students to watch a video on the structure and function of chromosomes and then answer a set of questions. This assignment received a low score on literacy challenge because there was no written text for students to engage with and the visual text presented to students (the video) was not addressed in instruction as a text (i.e., with a predictable organizational format), nor were their opportunities for students to reflect on their understanding of the "text."

Support for Literacy Challenge. The purpose of this dimension was to describe the degree to which a teacher supported the literacy challenge targeted by this assignment. Specifically, this dimension considers the degree of support students received for engaging meaningfully with biology texts (e.g., through opportunities to read in class and use targeted reading comprehension strategies) and for utilizing metacognitive processes to reflect on and assess their reading, thinking processes, and understanding of science concepts (for the purpose of increasing their understanding).

Rubric. Each score point is based on two criteria: Support for Engagement with Science Text and Support for Metacognitive Challenge. To receive a high score on this dimension, the assignment had to fully support the aspects of the literacy challenge dimension that were present in the assignment. Support for Engagement with Science Text considers whether the teacher provided current or previous teaching of reading comprehension strategies, structures the assignment around explicit steps (e.g., literacy routines), provided opportunities to read and work collaboratively in class, and made material and social resources available to students. Support for Metacognitive Challenge considers whether the teacher provided current or previous instruction in metacognitive processes, provided opportunities to practice using metacognition, structured the assignment to allow for collaborative meaning making, and provided material resources for students.

An example of an assignment with a high literacy challenge that received a high score for support asked students to draw and label detailed pictures that were analogies of specific biology standards. The support included teacher modeling, explicit lecturing, iterative literacy routines, and scaffolding in the form of work conferences and note taking. Although the literacy challenge did not include explicit metacognition to be supported, the literacy support for the existing text engagement was high. Conversely, in a lab assignment with high text complexity that received a low score for this support dimension, the teacher stated "[a]long with each presentation there are worksheets for the students. The worksheets include space for notes and any drawings, exercises the students need to do….For these biology classes I do not have differentiated levels of support." In this example, students were given a format for recording their answers, but they were not given support or resources to help find or think through their answers (e.g., peer discussion).

Quality of Evaluation Criteria. The purpose of this dimension was to assess the quality of the literacy evaluation criteria for the biology assignment in terms of their specificity, elaborateness, and potential for helping students improve their literacy skills. How clearly each aspect of the literacy evaluation criteria was defined is considered in the rating, as well as how much detail was provided for each of the criteria.

Rubric. Each score point was based on three criteria: Specificity, Information to Students, and Fidelity of Evaluation Criteria. Specificity considers whether a teacher's literacy evaluation criteria were clear, explicit, and elaborated. Information to Students considers whether the criteria were detailed and communicated to students to improve their literacy learning. Fidelity of Evaluation Criteria considers whether the criteria were reflected in the graded selection of student work samples.

For example, the following evaluation criteria for a presentation assignment received a high score: "There are two rubrics to [*sic*] this assignment: one rubric for the power point presentation and the second rubric for the oral presentation. The project (power point) [*sic*] rubric graded them on content, text elements, layout of slides, graphics, and writing mechanics. *See attached rubrics." In the PowerPoint rubric, the description of a content score of 4 states, "The content is written clearly and concisely with a logical progression of ideas and supporting information. It provides the audience with a sense of the project's main ideas. Information is accurate and current and is clearly supported by evidence." These evaluation criteria are very clear and specific and can help students improve their literacy learning. On the other hand, a teacher whose assignment received a low score for this dimension stated, "I simply read the essay, checking whether they accurately answered all of the questions in the prompt." In this instance, the evaluation criterion is vague and would not be helpful to students for improving their performance in literacy.

Alignment of Learning Goals and Task. This dimension was focused on the degree to which a teacher's stated literacy learning goals are reflected in the design of the biology assignment tasks students are asked to complete. Specifically, this dimension attempts to capture how well the biology assignment appears to promote the achievement of the teacher's literacy goals for student learning.

Rubric. Each score point was rated on the single criterion of alignment between goals and task. An assignment with a high score in this dimension was considered to have excellent alignment, with the task fully supporting the goals, neither calling for something not included in the other. On the other hand, an assignment receiving a low score in this dimension had very little alignment between goals and task, which resulted in the task providing little support for the assignment goals.

A hypothetical example of an assignment that would receive a high score on this dimension would have literacy learning goals that, for example, included learning to do Internet research and demonstrating understanding of that research using appropriate academic language. Good task alignment with the goals could be established by requiring students to conduct Internet research on a body system's structure and function and then prepare a PowerPoint presentation based on that research, including both text and images. In contrast, an assignment that would receive a low score on this dimension had as a literacy learning goal learning to evaluate the quality and reliability of internet resources for research purposes but did not have task dimensions that included any requirement that students demonstrate evaluation of these resources or present any criteria for why they selected one resource instead of another. Instead, the task only required students to write up their research based on the Internet sources. In this case, the stated learning goals were more expansive than those supported by the task.

Alignment of Learning Goals and Evaluation Criteria. This dimension was intended to describe the degree to which a teacher's literacy evaluation criteria supported the learning goals (i.e., the degree to which a teacher assessed students on the literacy skills and concepts they were intended to learn through the completion of the biology assignment).

Rubric. This dimension was scored on the single criterion of whether there is alignment between the teacher's stated literacy learning goals for students and the stated evaluation criteria.

An assignment that would receive a high score on this dimension would have learning goals that pertained to students' participation in high-quality metacognitive engagement with text and evaluation criteria that measured levels of participation and the quality of metacognitive engagement. An example of an assignment receiving a low score in this dimension is one in which the goals and evaluation criteria related to different aspects of the assignment. In an assignment about the process of inquiry in the field of science, the goals were written as "[t]o show the students that science is a constantly changing field," whereas the evaluation criteria were instead about the correctness of the writing form students were to produce to demonstrate their understanding: "All assignments are to be written in complete sentences or no credit is given."

Overall Quality of Assignment. This dimension was intended to provide a holistic rating of the quality of the biology assignment based on its level of cognitive challenge in terms of literacy demand, the support of the goals for student learning in literacy, the quality of the goals for student learning in literacy, the quality of the evaluation criteria, the alignment of the learning goals with the assignment task, and the alignment of the learning goals with the evaluation criteria.

Rubric. This dimension was rated on the single criterion of the collective level of quality of cognitive challenge, clarity, support, and application of learning goals and evaluation criteria in literacy.

Assignment Coversheet. In our request for assignments from teachers, we asked for a variety of materials, including six graded samples of student work, a completed coversheet describing the nature of the assignment (see Appendix G), and any other materials that could help us further understand the assignment, especially those that were handed out to students (assignment directions, rubrics, etc.). To facilitate the measurement of dimensions in the rubric, the coversheet asked teachers to respond to a series of questions related to the same rubric constructs, such as:

> What were your learning goals for this assignment? Please describe the science skills (and literacy skills, if applicable), biology concepts (and literacy concepts, if applicable), and/or facts you wanted students to learn as a result of completing this assignment. (Quality of the Goals for Student Learning)
>
> If applicable, please write the reading comprehension strategies students used to access the biology content for the genetics assignment text, e.g., taking notes, completing reading logs, engaging in "talking to the text," participating in discussions, or creating graphic organizers. Additionally, if possible, please attach a copy of those reading comprehension tools. (Literacy Challenge)
>
> Describe the support you provided for students in their reading and writing processes (e.g., scaffolding, modeling, explicit instruction, resources, discussion opportunities, etc.). If you differentiated levels of support for different students, please also describe this below. Please attach any support materials you provided to students. (Support for Literacy Challenge)

We also asked a variety of other questions to help us measure the rubric dimensions of Cognitive Challenge, Support for Cognitive Challenge, Quality of Evaluation Criteria, Alignment of Learning Goals and Task, and Alignment of Learning Goals and Evaluation Criteria.

Observations

As part of the validation work on the teacher assignments, we also conducted classroom observations with a subset of participants. For this purpose, we used an observation protocol that measured four dimensions: Instructional Strategies, Science Instruction, Literacy Content, and Classroom Culture.

A total of 14 participating teachers were observed in classrooms for 2 consecutive days (or two consecutive periods, in schools with block schedules). In order to examine general classroom processes in high school biology lessons. Observations were conducted using an observation protocol modified from the *Looking Inside the Class* observation protocol developed by Horizon Research, Inc. (HRI; Weiss, Banilower, McMahon, & Smith, 2001). Originally designed to investigate the current status and quality of science and mathematics instruction in the United States, the HRI observation protocol has also been used to identify factors that distinguish effective lessons from ineffective ones. The HRI observation protocol[7] was adapted to address the research goals of the current study by evaluating high school biology lessons in four dimensions (See Appendix H for the adapted observation protocol). Each of the dimensions are explained below.

The observation protocol begins with teachers' demographic information, such as gender, ethnicity, and number of years of teaching, as well as basic descriptive information about students, including their performance level compared to the student population of the school. Observers were also asked to describe the physical environment of the classroom as well as instructional materials used for the lesson.

Four major constructs were designed to measure the effectiveness of the observed lesson: instructional strategies, science content, literacy content, and classroom culture. Each of the constructs was measured using between five and eight indicators that reflect best instructional practices in science. Observers were asked to rate these indicators on a 7-point scale. The 1–5 range represented 1 (*not at all*) to 5 (*to a great extent*). A score of 6, representing (*don't know*), was reserved for instances in which there are not enough resources available to make a judgment, and 7, (*N/A*), indicated that the rating was not applicable. Additionally, observers were asked to provide a general rating for each construct with detailed supporting evidence.

[7] http://www.horizon-research.com/instruments/clas/cop.php

Observation Protocol Dimensions

Instructional strategies. Seven indicators were utilized to measure the quality of general instructional practices. Specifically, the aim of this construct was to evaluate how well the overall instruction was aligned with the notion of investigative science and how various aspects of the observed lesson were influenced by teachers' instructional expertise, classroom management strategies, assessment strategies, and questioning strategies. It also considered how instructional practices promoted student autonomy and how students' ability levels and their instructional needs were reflected in the delivery of the lesson.

Science instruction. This construct was aimed at measuring how science, as a field, is conveyed to students and how well science-specific instructional practices are implemented in the lesson. Five indicators were included to evaluate how science tends to be conveyed to students and how effectively elements of abstraction (e.g., theory building) are incorporated. Also considered is the appropriateness of instructional support for the learning of science (e.g., how the sense-making of science content occurs, how science content is applied to other areas and contexts, and how objects are used to mediate science learning).

Literacy content. This dimension was created for this study to address to what extent literacy is utilized and developed in content instruction. A total of eight indicators were developed to evaluate this variable. All of the items included in this dimension were targeted at rating the level of literacy engagement as a vehicle for learning. Specifically considered are the effectiveness of support for literacy development in concert with science instruction and the frequency of opportunities to build scientific literacy via reading, writing, and verbal scaffolding. Additionally, the dimension included an item to evaluate the quality of linguistic support for ELs and students with special needs.

Classroom culture. Five indicators were included to evaluate the quality of the learning environment from an effective perspective. Specifically, how effective the learning environment was, in terms of how student participation was promoted in various forms (e.g., participation through teacher-student interaction, student-student interaction, etc.), was considered. In addition to lesson ratings, observers were also asked to provide further details about the use of instructional time and lesson arrangement. This information was elicited in the last section of the observation protocol.

Observation Procedure and Ratings

To achieve high inter-reliability among researchers, the research team participated in an observer calibration session that involved a practice observation at one of the pilot-testing sites. The session also involved discussions around key constructs and individual indicators for lesson rating. During actual site visits, a lesson was observed by a pair of observers. After filling out an observation protocol separately, observers achieved consensus on each rating through a thorough discussion of observed classroom practices. All the information obtained from observations (including ethnographic field notes) was inputted electronically for analysis.

Analysis

We used simple correlations, regression analysis, and hierarchical linear models (HLM) to address the two research questions. First, we conducted reliability analysis on the ILA scores and also on the teacher assignment ratings. In order to check for reliability of the ILA scores, we conducted a generalizability study (G-study) with all trained raters on 30 randomly selected student assessments for each of the three rubric dimensions. G-studies provided a closer look at the different sources of error (variability) in the scores and their relative importance, as well as the overall reliability of the scores. For the reliability of teacher assignment ratings across two raters, kappa statistics were computed to evaluate the overall level of agreement and whether agreement among the raters exceeded chance levels.

In order to address our first research question, whether it is possible to differentiate the various cognitive demands associated with successful completion of science explanation tasks, we examined how each of the three aspects of student performance—Holistic, Content, and Language scores—relates to each other and to external measures of similar constructs (CST-Science and CST-English Language Arts). As a first step, correlations and regression analyses were conducted to examine the relationships among the three ILA scores, CST-Science, and CST-ELA. Then, HLM analyses were conducted to examine what proportion of the Content score versus the Language score contributed to the Holistic score. HLM analyses also provided information about the differential impact of EL status on the three scores. We hypothesized that if the scores represent separate dimensions as hypothesized then we will see performance differences between ELs and non-ELs on the holistic and language dimensions but not on the content dimension because it is supposed to be measuring content only.

Two sets of analyses were conducted, one for genetics and one for physiology. Given the limited sample size and the similarity in the constructs measured, we combined the Genetics 1 and Genetics 2 scores to represent the overall genetics sample. The same was done for the overall physiology sample.

In order to address the second research question, HLM analysis was conducted to compare whether the ILA was more sensitive to instructional differences than CST. If ILA is more instructionally sensitive then instructional differences would be more closely associated with student performance on ILA compared to CST. We will compare how significantly the instructional variables are related to both outcomes: ILA and CST.

HLM Analysis

The factors influencing student performance occurred in the context of classrooms, which gave rise to multilevel data. Usually, students within the same classroom are affected by similar factors, such as teacher characteristics, educational resources, and the environment of the classroom. Using hierarchical models provided a systematic way to investigate how teachers, and specifically the teacher assignment ratings, influenced student outcomes. Given the small sample size at Level 2, we included the teacher assignment ratings for each dimension separately. Variable names and descriptions are presented in Tables 6 and 7.

The final HLM model specified in our study is as follows:

Conditional Model - Algebra

$Y_{ij} = \beta_{0j} + \beta_{1j}$*(**EL STATUS**) + β_{2j}*(**GENDER**) + β_{3j}*(**MINORITY STATUS**) + r_{ij}

Where

r_{ij} = the deviation of student i from the average ILA or CST score for teacher j;

$\beta_{0j} = \gamma_{00} + \gamma_{01}$*(***TEACHER ASSIGNMENT DIMENSION RATING)*** + u_{0j}

$\beta_{1j} = \gamma_{10} + u_{1j}$

$\beta_{2j} = \gamma_{20} + u_{2j}$

$\beta_{3j} = \gamma_{30} + u_{3j}$

Where

β_{0j} = the average score on ILA or CST for teacher j;

β_{1j} = the EL status regression slope for teacher j;

β_{2j} = the Gender regression slope for teacher j; and

β_{3j} = the Minority regression slope for teacher j.

γ_{01} = the main effect of teacher assignment rating.

u_{0j} = the unique increment to the intercept associated with teacher j;

u_{1j} = the unique increment to the EL status slope associated with teacher j;

u_{2j} = the unique increment to the Gender slope associated with teacher j; and

u_{3j} = the unique increment to the Minority Status slope associated with teacher j.

Table 6

Description of Student-Level Indicators

Name	Description	Coding
Gender	Student gender	0 – Male
		1 – Female
EL	Student EL status	0 – Non-English learner
		1 – English learner
Grade	Grade level	(Grades 8, 9, 10, & 11)
CSTSC	Score on CST-Science test	(California state standardized test)
CSTELA	Score on CST-English Language Arts test	(California state standardized test)
HolG / HolP	ILA Holistic score for Genetics/Physiology	1 – 4
ConG / ConP	ILA Content score for Genetics/Physiology	1 – 4
LanG / LanP	ILA Language score for Genetics/Physiology	1 – 4

Table 7

Description of Teacher-Level Indicators

Name	Description	Coding
Exp	Teacher's years of experience teaching	Continuous
CG_GO	Teacher's Quality of Goals score on Content teacher assignment rating for the Genetics assignment submitted	[1-4]
CG_CH	Teacher's Cognitive Challenge score on Content teacher assignment rating for the Genetics assignment submitted	[1-4]
CG_SU	Teacher's Support for Cognitive Challenge score on Content teacher assignment rating for the Genetics assignment submitted	[1-4]
CG_EV	Teacher's Evaluation Criteria score on Content teacher assignment rating for the Genetics assignment submitted	[1-4]
LG_GO	Teacher's Quality of Goals score on Language teacher assignment rating for the Genetics assignment submitted	[1-4]
LG_CH	Teacher's Cognitive Challenge score on Language teacher assignment rating for the Genetics assignment submitted	[1-4]
LG_SU	Teacher's Support for Cognitive Challenge score on Language teacher assignment rating for the Genetics assignment submitted	[1-4]
LG_EV	Teacher's Evaluation Criteria score on Language teacher assignment rating for the Genetics assignment submitted	[1-4]
CP_GO	Teacher's Quality of Goals score on Content teacher assignment rating for the Physiology assignment submitted	[1-4]
CP_CH	Teacher's Cognitive Challenge score on Content teacher assignment rating for the Physiology assignment submitted	[1-4]
CP_SU	Teacher's Support for Cognitive Challenge score on Content teacher assignment rating for the Physiology assignment submitted	[1-4]
CP_EV	Teacher's Evaluation Criteria score on Content teacher assignment rating for the Physiology assignment submitted	[1-4]
LP_GO	Teacher's Quality of Goals score on Language teacher assignment rating for the Physiology assignment submitted	[1-4]
LP_CH	Teacher's Cognitive Challenge score on Language teacher assignment rating for the Physiology assignment submitted	[1-4]
LP_SU	Teacher's Support for Cognitive Challenge score on Language teacher assignment rating for the Physiology assignment submitted	[1-4]
LP_EV	Teacher's Evaluation Criteria score on Language teacher assignment rating for the Physiology assignment submitted	[1-4]

For each of the two topic areas, we examined the Holistic, Content, and Language scores as outcome variables separately. CST-Science score was also treated as outcome variable to compare the sensitivity of the various outcome measures in detecting instructional differences. We had a total of 16 teacher assignment ratings across both content and literacy dimensions in genetic and physiology for each teacher. Due to small sample size at the teacher level, we created new teacher assignment variables by combining the ratings from several dimensions together. Also, we included gender, EL status, and minority status in Level 1 on all the analysis.

Table 8

HM Model Description

Subject	Outcomes	Teacher variables included besides teacher experience	Student variables included
Genetics	Holistic Content Language CST-Sc	1. Content Clarity of Goals (CG_GO)	EL, Gender, Minority, Grade level
		2. Content Cognitive Challenge (CG_CH)	EL, Gender, Minority, Grade level
		3. Content Cognitive Support (CG_SU)	EL, Gender, Minority, Grade level
		4. Content Evaluation Criteria (CG_EV)	EL, Gender, Minority, Grade level
		5. Language Clarity of Goals (LG_GO)	EL, Gender, Minority, Grade level
		6. Language Cognitive Challenge (LG_CH)	EL, Gender, Minority, Grade level
		7. Language Support (LG_SU)	EL, Gender, Minority, Grade level
		8. Language Evaluation Criteria (LG_EV)	EL, Gender, Minority, Grade level
Physiology	Holistic Content Language CST-Sc	1. Content Clarity of Goals (CP_GO)	EL, Gender, Minority, Grade level
		2. Content Cognitive Challenge (CP_CH)	EL, Gender, Minority, Grade level
		3. Content Cognitive Support (CP_SU)	EL, Gender, Minority, Grade level
		4. Content Evaluation Criteria (CP_EV)	EL, Gender, Minority, Grade level
		5. Language Clarity of Goals (LP_GO)	EL, Gender, Minority, Grade level
		6. Language Cognitive Challenge (LP_CH)	EL, Gender, Minority, Grade level
		7. Language Support (LP_SU)	EL, Gender, Minority, Grade level
		8. Language Evaluation Criteria (LP_EV)	EL, Gender, Minority, Grade level

CHAPTER 3: RESULTS AND CONCLUSIONS

This chapter is organized into four sections. The first section presents the reliability information on the ILA and the teacher assignment ratings. The next two sections present the results corresponding to our two main research questions. The overall findings and discussion of the results are also included at the end of this chapter.

Reliability of the Instruments

Reliability of ILA Scores

To examine the reliability of ILA scores, we investigated specifically the amount of variability in scoring due to raters. The results from the G-studies suggest that the rater reliability and the overall ILA scores were generally high.

The generalizability coefficients were .955, .944, and .956, for Holistic, Content, and Language, respectively. Table 9 below provides the variance components and the generalizability coefficients for each dimension.

The vast majority of the variation in students' scores, in all three cases (Holistic, Content, and Language), was due to variation in the student papers themselves (68.2%, 62.9%, and 68.3%, respectively) and to the combined effect of the error (23.1%, 32.4%, and 25.8%, respectively). Variation in scores due to rater inconsistency was low for each category: differences among the raters accounted for only 8.7% of the total variability in Holistic scores, 4.7% of the total variability in Content scores, and 5.8% of the total variability in Language scores. The results of the generalizability theory analysis suggest that, on the whole, the raters were highly consistent and reliable. Upon closer examination of the reliability papers, we found that the scores of Rater 7 frequently deviated from the other raters' scores. Thus, student responses scored by Rater 7 were re-scored by other raters.

Table 9

G-study Summary: 10 raters, 30 papers.

Component	Holistic	Content	Language
Var(rater)	.08 (8.7%)	.04 (4.7%)	.05 (5.8%)
Var(paper)	.62 (68.2%)	.47 (62.9%)	.60 (68.3%)
Var(paper*rater, error)	.21 (23.1%)	.24 (32.4%)	.23 (25.8%)
G coefficient	.96	.94	.96

Note. * = Interaction term between paper and rater.

Teacher Assignment Reliability

Each assignment was scored by two raters (total $n = 4$) on 14 dimensions. Through discussion and using initial, independent scores as a focus for these discussions, the raters then established final consensus scores for all dimensions on both rubrics for each assignment. For the HLM analysis, only 8 of the 14 dimensions were included in the final analysis due to multicollinearity of the scores. As shown in Table 10, agreement among raters before consensus ranged from 42% to 61%. However, all ratings included in the analyses are based on the consensus scores.

Table 10

Rater Agreement

Rater	Exact agreement	Kappa	Std. Err.	Prob
1 & 2	0.42	0.16	0.09	0.03
1 & 3	0.46	0.25	0.12	0.02
2 & 3	0.49	0.28	0.05	0.00
2 & 4	0.61	0.44	0.04	0.00
3 & 4	0.54	0.36	0.09	0.00

Before we present the results from the HLM, we first briefly describe the results from the teacher assignments including the descriptions of the assignments and the ratings, which were included as key indicators of instructional practice in the HLM analyses.

Teacher Assignment Ratings

We found that the majority of teachers' assignments followed a similar pattern of classroom activities. In general, the structure of activities in the collected assignments included: (a) an input phase, in which students learned new information; (b) an activity phase, in which students applied this information to a specific context; (c) a demonstration-of-knowledge phase, in which students wrote up, illustrated, or orally presented their understandings; and (d) an assessment phase, in which teachers reviewed and graded student work. Not all assignments included all four phases, though many did.

Of the 46 assignments reviewed in this study, the majority included the input of new information to students through the reading of text (89%; see Table 11). Fifty-four percent of the assignments included reading from the course textbook only, whereas 22% included reading from the textbook as well as alternate sources such as the Internet, and 11% included reading only alternate sources (e.g., Internet, journal articles, magazines, and newspapers).

An assignment including reading also usually included teacher input in the form of a lecture and/or the results of an inquiry or problem-solving situation (e.g., a lab conducted by students).

Students' learning processes were frequently supported with a variety of scaffolding techniques, most notably, note-taking during the input of new information (63%) and discussion opportunities (46%). Frequently, teachers who provided these scaffolds for student learning processes incorporated more than one within their assignments (e.g., including discussion opportunities and requiring the use of note taking and graphic organizers within the context of a single assignment).

Teachers also provided support for students (80%) through various instructional styles (80%). Most frequently, this support took the form of providing students with information handouts and example student work (37%), using visual aids such as PowerPoint presentations (28%), providing written feedback to students (28%), and modeling (22%). Nearly half of the tasks focused on the process of explaining or describing a biological process or concept (48%). Other assignments included a lab structure with an inquiry or problem-solving component (37%) and/or a research component (22%). Three of the assignments in the study included a combination of lab, research, and explanation task components.

In over half of the assignments (54%), students were asked to demonstrate their knowledge in multiple ways. For example, one assignment on genetic disorders asked students to research a genetic disorder of their choice and then demonstrate their knowledge through the creation of a pamphlet about the genetic disorder and an oral PowerPoint presentation. As the primary means of assessment, most teachers asked students to demonstrate their knowledge through writing a paragraph or more (70%). This writing was generally in the form of an essay or report. Teachers also asked students to visually represent their understanding through illustrations or diagrams (28%), fill out worksheets (30%), and/or provide short-answer responses to questions (35%). This work was mostly carried out individually (83%), though sometimes in pairs (11%) or in small groups (24%). Usually the group work was in the context of a lab experiment and was followed by individual student write-ups of the work (therefore a single assignment could include more than one grouping configuration). In terms of assessment strategies, more than half of the teachers did not appear to provide students written information about their evaluation criteria before the completion of the assignment. Only 41% of teachers used grading rubrics. These frequently appear to have been given to students prior to completion of the assignment and include specific information about the strengths and weaknesses of individual student work.

The majority of genetics assignments were related to Mendelian genetics and consisted of students determining the likelihood of offspring having a given trait (39% of all assignments). Other genetics assignments pertained to the structure of DNA itself (e.g., amino acid sequencing) and the process of genetic engineering (15% of all assignments). Seventeen percent of assignments were cell biology assignments and focused on the structure and function of cells. The physiology assignments were nearly all related to body systems—immune, circulatory, nervous, respiratory, endocrine (26% of all assignments)—and only one physiology assignment was related to ethical issues concerning HIV testing. The following table provides more description in the frequencies of various aspects of the collected assignments.

Table 11

Assignment Characteristics and Frequencies

Assignment characteristics	*N*	%
Input		
Reading (written text)	41	89
Lecture (generally paired with a reading activity)	26	57
Experiential: inquiry and/or problem-solving situation (generally includes some reading)	16	35
Watching film/video	7	15
Student learning support activities		
Discussion	21	46
Note taking	29	63
Graphic organizers	12	26
Reading comprehension strategies	10	22
Teacher instructional support		
Models	10	22
Uses visual aids including PowerPoint, overheads, etc.	13	28
Provides material resources, (e.g., information handouts and work samples)	17	37
Provides feedback to students	13	28
Other; conferencing and clarifying concepts during read aloud	3	7
Uses explicit instruction	4	9
Task context (assignments could include more than one of the following processes)		
Lab structure (e.g., including inquiry/problem solving phase and write up)	17	37
Research	10	22
Written and/or illustrated explanations and descriptions	22	48
Demonstration of knowledge		
Writing: up to a few sentences for each section (e.g., answers to questions)	16	35
Writing: paragraph or more (usually an essay or report)	32	70
Oral presentations: including PowerPoint	7	15
Illustration/cut and paste/ diagram	13	28
Worksheet	14	30
Other: making video and model	2	4

(table continues)

Table 11 (continued)

Assignment Characteristics and Frequencies

Assignment characteristics	*N*	%
Grouping configurations		
Individual	38	83
Pair	5	11
Small group (usually with individual work following)	11	24
Assessment		
Uses a rubric	19	41
Topic		
Mendel's Law (e.g., genetic disease, heredity)	18	39
DNA structure and Genetic Engineering (e.g., amino acid sequence, gene splicing)	7	15
Cell Biology (e.g., cell structure and function)	8	17
Body systems (i.e., immune, circulatory, nervous, respiratory, endocrine)	12	26
Other (ethics for testing HIV)	1	2

Assignment Ratings Analysis

Overall, the assignments did not receive high scores (a score of 3 considered adequate teacher practice), with only 19% of teachers receiving a 3 or 4 on the Overall Content quality dimension and 2% of teachers receiving 3 or 4 on the Overall Literacy quality dimension.[8] In general, the Content dimensions scored higher than the Literacy dimensions by a full score point, with assignments scoring on average between 2 and 3 on all Content dimensions and between 1 and 2 on all Literacy dimensions. Considering the dimensions separately, most Content dimensions received relatively high scores, with over 50% of teachers demonstrating adequate to excellent teacher practice (i.e., receiving a score of 3 or 4) in 4 out of the 7 dimensions. These dimensions were: Quality of Content Goals, Cognitive Challenge, Support for Cognitive Challenge, and Alignment of Goals and Task. Within the Literacy dimensions, the highest scoring dimension was Alignment of Goals and Task, with 15% of teachers scoring adequately, and the lowest scoring was Literacy Challenge, with 0% of teachers scoring adequately. This is not surprising, because high school biology teachers frequently

[8] These numbers relate to the percentage of teachers receiving these scores on either one or both of their submitted assignments (each teacher submitted two).

expect their students to read and write science text as part of their classroom activities but consider instruction in these areas (e.g., of reading comprehension strategies) to be the purview of English Language Arts teachers (Hays & Harris, 2003; Morrow, Pressley, Smith, & Smith, 1997).

Table 12

Percentage of Teacher Assignment Scores by Dimension

	Genetics		Physiology		All	
Dimension	% 1+2	% 3+4	% 1+2	% 3+4	% 1+2	% 3+4
Content						
Quality of Goals	44	56	43	57	43	57
Cognitive Challenge	32	68	57	43	43	57
Support for Cognitive Challenge	52	48	43	57	48	52
Quality of Evaluation Criteria	64	36	71	29	67	33
Alignment of Goals and Task	36	64	38	62	37	63
Alignment of Goals and Criteria	68	32	81	19	74	26
Overall	76	24	81	19	78	22
Literacy						
Quality of Goals	96	4	9	1	93	7
Literacy Challenge	1	0	1	0	1	0
Support for Literacy Challenge	92	8	95	5	93	7
Quality of Evaluation Criteria	84	16	95	5	89	11
Alignment of Goals and Task	84	16	86	14	85	15
Alignment of Goals and Criteria	88	12	1	0	93	7
Overall	96	4	1	0	98	2

Genetics and physiology assignments were generally of equal quality, as shown above in Table 12. Twenty-four percent of genetics assignments and 22% of physiology assignments received a 3 or better on the Overall Content quality dimension. Four percent of genetics and 0% of physiology assignments received a 3 or better on the Overall quality dimension for Literacy. Comparing the specific Content dimension scores in genetics and physiology, there is general consistency in teacher performance in most dimensions.

However, in the Content Cognitive Challenge and Alignment of Goals & Evaluation Criteria dimensions, teachers scored considerably lower in Physiology than in Genetics (with 19% and 13% differences, respectively). The low level of the scores in these dimensions in physiology could be the result of physiology receiving less instructional coverage than genetics and, as such, resulting in lower teacher expectations. Many teachers in the study commented that physiology was the last subject they covered during the school year and that they were pressed for time during this instructional period. Interestingly, scores on Content Support for Cognitive Challenge remain high in physiology even though scores on Cognitive Challenge are low. This dynamic of high support for low cognitive challenge is similar to our findings in the earlier CRESST study (Aguirre-Muñoz et al., 2005), where we saw teachers overly scaffolding assignments in a manner that reduced the cognitive demand. This was evident in cases where teachers deemed their students' capacity to perform well on the assignment as limited. In the context of this current study, it may also be the case that although teachers continued to set goals based on the standards, they adjusted their evaluation criteria to match their expectations of student performance. This may account for the low alignment score in physiology between goals and evaluation criteria. Genetics, in contrast, generally received substantial instructional time during the school year among our sample of teachers, as it figures highly in the CST Biology exam.[9] Among the Literacy dimensions for physiology, Quality of Evaluation Criteria and Alignment of Goals & Evaluation Criteria also scored significantly lower than their genetics counterparts. This may be due to the same reasons mentioned above; that teachers' evaluation criteria in physiology may be based on lower expectations of student performance.

The teacher ratings from each of these dimensions were included in the ordinal logistic regression to examine the relative significance of these factors in explaining the student performance on the ILA. The descriptive information on the teacher assignment ratings for genetics and physiology assignments are presented in Table 13 and Table 14, respectively.

[9] Although physiology is also covered in the CST Biology exam, many teachers commented that students learned physiology in 7th grade and so they felt it was less important to dedicate their limited instructional time to the subject, as compared to other newly introduced subject such as genetics.

Table 13

Teacher Descriptives – Genetics

Teacher variables	*N*	Minimum	Maximum	Mean	Std. deviation
Teacher years of experience	11	1	34	11.09	9.46
Teacher assignment ratings					
Content					
Quality of Goals	11	1	4	2.73	0.90
Cognitive Challenge	11	1	4	3.00	0.77
Support for Cognitive Challenge	11	2	4	2.64	0.81
Quality of Evaluation Criteria	11	1	3	2.09	0.54
Literacy					
Quality of Goals	11	1	2	1.27	0.47
Cognitive Challenge	11	1	2	1.36	0.50
Support for Cognitive Challenge	11	1	3	1.36	0.67
Quality of Evaluation Criteria	11	1	2	1.27	0.47

Table 14

Teacher Descriptives – Physiology

Teacher variables	*N*	Minimum	Maximum	Mean	Std. deviation
Teacher years of experience	8	3	15	7.06	3.86
Teacher assignment ratings					
Content					
Quality of Goals	8	2	4	2.75	0.71
Cognitive Challenge	8	2	3	2.25	0.46
Support for Cognitive Challenge	8	2	4	2.38	0.74
Quality of Evaluation Criteria	8	2	4	2.75	0.89
Literacy					
Quality of Goals	8	1	3	1.50	0.76
Cognitive Challenge	8	1	2	1.38	0.52
Support for Cognitive Challenge	8	1	2	1.38	0.52
Quality of Evaluation Criteria	8	1	4	1.50	1.07

Analysis of Observation Protocol Ratings in Relation to Teacher Assignment Ratings

As part of the validation work on the teacher assignments, we examined the observation data in relation to the teacher assignment ratings. In the observation protocol, we included sets of items that are aligned with teacher assignment rating dimensions. Specifically, we included items in the observation protocol aligned with 2 dimensions from Content rubric and 2 dimensions from the Literacy rubric. These dimensions are Cognitive Challenge[10], Support for Cognitive Challenge[11], Literacy Challenge[12], and Support for Literacy Challenge[13]. In the observations, we found that overall teachers were rated higher in the Content dimensions—Cognitive Challenge and Support for Cognitive Challenge, than in the corresponding Literacy Dimensions. This finding is consistent with Teacher Assignment ratings in which teachers generally scored higher in the Content dimensions compared to Literacy dimensions as well.

For comparison purposes, we defined "good" instructional practice in a particular dimension as a score of 4 or higher using the observation protocol. For example, an observed lesson that received a 4 in the Support for Cognitive Challenge included the teacher providing symbolic representations of ideas through graphs and pictorial images, concepts given "real world" examples, and students given the opportunity to discuss answers to teacher questions in small groups before sharing with the whole class. In the teacher assignment ratings that had a 1–4 point scale, we defined "good" teacher practice on an instructional dimension if it received a score of 3–4. For example, in a teacher assignment that received a score of 3 in Support for Cognitive Challenge, students were provided with work conferences, lectures, note-taking opportunities, teacher modeling, and storyboarding to help them with the assignment. These definitions of good instructional practice are based on the rubrics and score point definitions for the two instruments. As shown in Table 15, the percentage of teachers demonstrating good instructional practices varied between Content and Literacy for both instruments.

[10] Observation protocol items measuring Cognitive Challenge are I.A.1, I.A.4, II.A.4, and IV.A.5
[11] Observation protocol items measuring Support for Cognitive Challenge are I.A.5, IV.A.1, IV.A.3, and IV.A.5
[12] Observation protocol items measuring Literacy Challenge are II.A.2, III.A.1, III.A.3, and III.A.5
[13] Observation protocol items measuring Support for Literacy Challenge are III.A.2 and III.A.6

Table 15

Percentage of Observed Teachers Who Provided Good Instructional Practice in Four Instructional Dimensions

Instrument	Cognitive Challenge	Cognitive Challenge	Literacy Challenge	Literacy Support
Teacher assignment ratings	77	46	0	8
Observation protocol	31	46	8	23

When comparing the percentage of teachers with good instructional practices in relation to the different dimensions, we see that the measurement of Cognitive Challenge in the two instruments is substantially different (71% and 31%) whereas the measurement of Cognitive Support is the same across the two instruments (46%). The percentage of teachers demonstrating good literacy practices as measured in the two instruments is fairly consistent across the two instruments, particularly in the Literacy Challenge dimension with 0% with teacher assignments and 8% with observation protocol.

Overall, three of the measured dimensions demonstrated very similar findings; in particular, the findings of Cognitive Support were the same across both instruments. Although the Literacy Support dimensions were measured as demonstrating slightly higher quality practice in the Observation Protocol, they were still comparable with the Teacher Assignment Ratings. In comparison, the Cognitive Challenge dimension shows a significant difference between the two instruments. This difference could be a result of the teachers inflating reports of their assignment's cognitive challenge in their written descriptions, because they would be well aware that cognitive challenge is highly valued in science education. Alternately, this difference could be the result of using instrument that may have only provided a snap shot of an assignment spanning many days or weeks. Given the differences observed in the cognitive challenge dimension using these two instruments challenges the validity of this construct and thus, any analysis including this dimension will be interpreted with caution.

Based on the teacher assignment ratings and the student ILA scores, we were able to address the two main research questions posed in the current study. The findings corresponding to each research question are presented next.

Content Versus Language Demands

We used simple correlations, regression analysis, and hierarchical models (HM) to address the first research question: ***Is it possible to differentiate the various cognitive demands associated with the successful completion of science explanation tasks using the***

newly developed assessment rubric? Specifically, what proportion of content versus language skills contributes to the overall evaluation of student performance? The results are reported separately for each topic area.

Genetics

For genetics, as shown in Table 16 correlation analyses indicate that both Content (r = 0.87) and Language scores (r = 0.85) were highly correlated with the Holistic score. In comparison, the correlation between Content and Language was 0.76.

The correlations between each of the ILA scores and each of the CST scores ranged from 0.52 to 0.61. Interestingly, of the three ILA scores—Holistic, Content, and Language—it was the Language score that had the highest correlation with the CST, both CST-Science (CST_Sc) and CST-English Language Arts (CST_ELA).

Table 16

Correlations Among the Outcome Variables

Student Outcome	CST_Sc	CST_ELA	Holistic	Content	Language
CST_Sc	1				
CST_ELA	0.80	1			
Holistic	0.58	0.56	1		
Content	0.56	0.52	0.87	1	
Language	0.61	0.61	0.85	0.76	1

As an initial exploratory analysis and for the purpose of interpretation, we conducted a linear regression analysis to examine the relationship between the Holistic score and the other two ILA scores (Content and Language). According to the results of the regression analysis, about 84% of the Holistic score can be attributed to the Content and Language scores (R^2 = .84). Also, as shown in Table 17 below, based on the regression coefficients, Content and Language scores seem to be contributing about equally to the explanation of the Holistic score. In other words, as hypothesized, the results of the regression analysis suggest that Holistic scores are a representation of students' language skills as well as their content knowledge.

Table 17

Regression Results for Genetics (Outcome Variable = Holistic)

Variables	*B*	*T*	*Sig*
Content	.52	17.80	.00
Language	.43	15.50	.00

We also conducted HM analyses to check for validity of these three separate dimensions. We found that EL status was negatively associated with performance on both holistic (p = .02) and language dimensions (p = .00) but not on the content dimension (p = .21). Also, magnitude of difference in performance was slightly higher between EL and non-ELs on the language dimension (ß = -1.74) compared to holistic dimension (ß = -1.06). This result suggests that performance on content dimension is not related to the language proficiency and language skills but the performance on holistic and language are significantly related to language skills.

Physiology

As shown in Table 18, for physiology, the ILA Holistic score is highly correlated with both the Language score (r = 0.75) and the Content score (r = 0.72). The correlation between Content and Language scores for physiology was only 0.61. The correlations between each of the ILA scores and each of the CST scores ranged from 0.39 to 0.51. Of the three ILA scores, the Holistic score was the most highly correlated with the CST, both CST-Science (CST_Sc) and CST-English Language Arts (CST_ELA).

Table 18

Correlation Among the Outcome Variables

Student Outcome	CST_Sc	CST_ELA	Holistic	Content	Language
CST_Sc	1.00				
CST_ELA	0.79	1.00			
Holistic	0.51	0.54	1.00		
Content	0.43	0.45	0.81	1.00	
Language	0.49	0.54	0.82	0.71	1.00

We also conducted a linear regression analysis to examine the relationship between the Holistic score and the other two ILA scores (Content and Language) for physiology. Based on the analysis, about 77% of the Holistic score can be attributed to the Content and

Language scores (R^2 = .77). As shown in Table 19, as with Genetics, the regression coefficients illustrate that Content and Language scores seem to be contributing almost equally to the explanation of the Holistic score. Consistent with the genetics data, as hypothesized, the regression analysis results suggest that Holistic scores are a representation of both students' content knowledge as well as their language skills for physiology.

Table 19

Regression Results for Genetics

Variables	*B*	*T*	*Sig*
Content	.52	14.10	.00
Language	.47	14.81	.00

We also conducted HM analyses to check for validity of these three separate dimensions. For physiology, we found that EL status was negatively associated with performance on all three ILA dimensions.

Sensitivity of ILA Versus CST

To address our second research question: ***What is the validity of ILA as an outcome measure? Specifically, is ILA more sensitive to detecting instructional differences compared to the other standards-based assessment?***, we also used a 2-level ordinal logistic hierarchical model (HM) analysis to compare how well the ILA differentiates the instructional quality (using teacher assignment ratings as indicators) compared to CST. Student demographic information, teachers' years of experience, and teacher assignment ratings data were analyzed in concert with student performance results using a 2-level ordinal logistic HM.

Ordinal Logistic HLM Results

In general, the scores on the three dimensions of ILA were higher on genetics compared to physiology. The mean scores on genetics were 2.21 on holistic, 2.02 on content, and 2.37 on language. In comparison, for physiology, the means were 1.88 on holistic, 1.80 on content and 1.96 on language. Also, in both genetics and physiology, language scores were higher than the other two scores. However, these differences were negligible. Scores on reading comprehension questions were overall very high. The percentage of students answering 2 out 3 items correctly on the reading comprehension section were 78% for genetics and 73% for physiology. The ILA scores also varied across the grade levels with higher performance in lower grade levels. For most schools, regular biology courses are offered at Grade 9. However, more accelerated Grade 8 students are eligible to take the course as well as students in Grade 10 or Grade 11 who are not enrolled in honors or AP biology courses are also offered an option to take the regular biology course with Grade 9 students. Thus, the distribution of ILA scores that we observed may reflect the qualitative differences in students enrolled in these classes characterized by the different grade levels. Descriptive information about the ILA score distribution by student background characteristics is provided in Table 20 and

Table 21.

Table 20

Descriptive Information on the ILA by Key Background Variables – Genetics

Background variables	*N*	Mean score on ILA holistic	Mean score on ILA content	Mean score on ILA language
Gender				
Male	213	2.23 (0.79)	2.03 (0.82)	2.35 (0.84)
Female	252	2.20 (0.79)	2.01 (0.76)	2.39 (0.83)
Minority status				
Non-minority	247	2.46 (0.78)	2.22 (0.84)	2.60 (0.76)
Minority	218	1.94 (0.71)	1.80 (0.67)	2.10 (0.83)
EL status				
Non-EL	442	2.24 (0.83)	2.04 (0.78)	2.41 (0.82)
EL	19	1.68 (0.90)	1.66 (0.93)	1.53 (0.70)
Grade				
8	73	2.63 (0.71)	2.32 (0.80)	2.85 (0.66)
9	273	2.16 (0.75)	1.95 (0.72)	2.30 (0.81)
10	101	2.14 (0.89)	2.07 (0.92)	2.28 (0.90)
11	18	1.97 (0.79)	1.72 (0.75)	2.00 (0.73)

Table 21

Descriptive Information on the ILA by Key Background Variables – Physiology

Background variables	*N*	Mean score on ILA Holistic	Mean score on ILA Content	Mean score on ILA Language
Gender				
Male	209	1.88 (0.68)	1.81 (0.62)	1.93 (0.69)
Female	225	1.89 (0.61)	1.80 (0.53)	1.99 (0.66)
Minority status				
Non-minority	167	2.18 (0.63)	2.00 (0.58)	2.28 (0.65)
Minority	267	1.69 (0.58)	1.68 (0.54)	1.76 (0.61)
EL status				
Non-EL	391	1.91 (0.65)	1.83 (0.57)	2.01 (0.67)
EL	43	1.58 (0.54)	1.58 (0.54)	1.57 (0.57)
Grade				
8	14	2.19 (0.60)	1.81 (0.43)	2.19 (0.72)
9	278	1.99 (0.62)	1.91 (0.57)	2.09 (0.63)
10	124	1.60 (0.63)	1.55 (0.51)	1.69 (0.68)
11	18	1.82 (0.56)	1.79 (0.61)	1.68 (0.47)

Genetics

For genetics, we found that none of the teacher assignment ratings were associated with performance on the ILA. However, the level and support for cognitive challenge was negatively associated with student performance on the CST-Science. We may be able to attribute this negative relationship to the misalignment between the performance level of students and the level of expectations and types of support or scaffolding that the teacher provides to the students to accomplish the task. As previous studies (Aguirre-Muñoz et al., 2005) have shown scaffolding or extensive support without gradual move towards independence may limit opportunities for students to actively engage in and have input in their self-regulated learning which is considered a critical skill for scientific inquiry. Also, the high level of cognitive support may represent a mediating variable for lower ability students. The indication of high level of support may in fact be attributed to the needs of the lower ability students. Thus, the negative relationship between the levels of support and performance may be representative of the association between lower ability students and low performance rather than direct association between support and performance.

As shown below in Table 22, at the student level, as discussed previously, we found that EL status was negatively associated with performance on only holistic and language dimensions. EL status was not associated with performance on content ILA. These results suggest that EL status may be negatively impacting student's performance on holistic score although the student may have adequate content knowledge. This result provides a strong support for the rationale and also feasibility in differentiating different cognitive demands through explicit scoring criteria. Also, minority status was also negatively associated with scores on the holistic and content ILAs. The other student background variables including gender, grade levels, and reading comprehension scores were not associated with performance on the three ILAs.

For CST-Science, we found that EL status and grade level were negatively associated with performance. Given that biology is typically taken in 9th grade, we would expect students who are taking this course out of the normal course taking sequence to be qualitatively different from Grade 9 students.

Table 22

Effects of Teacher- and Student-Level Variables on ILA Genetics Scores

Variables	Holistic	Content	Language	CST
Teacher-level				
Teacher experience	NS	NS	NS	NS
Content				
Quality of Goals	NS	NS	NS	NS
Cognitive Challenge	NS	NS	NS	NS
Support for Cognitive Challenge	NS	NS	NS	**-34.86****
Quality of Evaluation Criteria	NS	NS	NS	NS
Literacy				
Quality of Goals	NS	NS	NS	NS
Literacy Challenge	NS	NS	NS	NS
Support for Literacy Challenge	NS	NS	NS	NS
Quality of Evaluation Criteria	NS	NS	NS	NS
Student-level				
Gender	NS	NS	NS	NS
EL status	**-0.32***	NS	**-0.61****	**-28.63****
Grade level	NS	NS	NS	**-17.63****
Minority status	**-0.28****	**-0.19***	NS	NS

Note. **p* < .05. ***p* < .01. ****p* < .001. NS = not statistically significant.

Physiology

Physiology results were similar to findings from genetics. We found that none of the teacher assignment rating variables had an impact on student performance on the three ILA dimensions.

As shown below in Table 23, at the student level, for physiology, EL status was negatively associated with performance on all three ILA dimensions and also CST. Also, minority status was associated with lower performance on the three ILAs. The other student background variables including gender and grade levels were not associated with performance on the three ILAs. The lack of differential performance on three ILA dimensions on physiology may be due to relatively lower content coverage for physiology, as compared to genetics. As mentioned previously, many teachers in the study commented that physiology was the last subject they covered during the school year and the emphasis on the content coverage is relatively weak for physiology compared to other topic areas. Thus, for students to perform well on physiology, they had to rely more on the text prompt in the assessment to acquire content knowledge, which required more language skills for comprehension, and knowledge of technical vocabulary compared to genetics text.

Again, consistent with genetics, we found that EL status was negatively associated with CST. For CST, we also found that grade level was negatively associated with student outcome. Given that biology is typically taken in 9th grade, we would expect students who are taking this course out of normal course taking sequence to be qualitatively different from Grade 9 students.

Again, it is difficult to determine whether this difference in performance is truly due to achievement gap in content understanding or the language demands confounded in the content assessment.

Table 23

Effects of Teacher- and Student-Level Variables on ILA Physiology Scores

Variables	Holistic	Content	Language	CST
Teacher-level				
Teacher experience	NS	NS	NS	NS
Content	NS	NS	NS	NS
Quality of Goals	NS	NS	NS	NS
Cognitive Challenge	NS	NS	NS	NS
Support for Cognitive Challenge	NS	NS	NS	NS
Quality of Evaluation Criteria	NS	NS	NS	NS
Literacy				
Quality of Goals	NS	NS	NS	NS
Literacy Challenge	NS	NS	NS	NS
Support for Literacy Challenge	NS	NS	NS	NS
Quality of Evaluation Criteria	NS	NS	NS	NS
Student-level				
Gender	NS	NS	NS	NS
EL status	-0.33**	**-0.39****	0.39**	-25.72***
Grade level	NS	NS	NS	-11.80**
Minority status	-0.31**	-0.28*	-0.28*	NS

Note. *$p < .05$. **$p < .01$. ***$p < .001$. NS = not statistically significant.

Summary of Findings and Discussion

Our recent review of content assessments revealed that language expectations and proficiencies are often implicitly embedded within the assessment criteria. Based on a review of performance assessments used in high school biology settings, we have found a recurring discrepancy between assessment scoring criteria and performance expectations. As mentioned previously, without explicit scoring criteria to evaluate the language performance, it is difficult to determine how much of the overall performance quality can be attributed to language skills versus content knowledge. This is an especially important validity question for ELs under the current state assessment mandates. Given this challenge, our current study was guided by two principal research questions: first, we wanted to examine to what extent content knowledge versus language skills contribute to the overall evaluation of student performance, and secondly, we were interested in examining the relative sensitivity of the standards-based assessment in detecting instructional impact which is an important component of the assessment goal compared to more proximal measure of student understanding using performance-based assessment.

Based on our analyses, we found that language skills are an important component of the overall evaluation of student performance on content-based assessments. In this study, we were able to show the feasibility of differentiating the various cognitive demands associated with content assessment. With both the language expectations and the evaluation criteria made explicit through our scoring rubric, we were able to show the significance of language proficiency in content assessment. In our pilot study, for genetics at least, we found differences only in performance between EL and non-ELs when they were evaluated holistically and in the language dimension but not in the content dimension. This finding suggests that the current evaluation system implicitly embeds language expectations into the scoring criteria and confounds content knowledge performance with language demands. This finding highlights the validity issues surrounding content assessments with ELs. The evaluation of current content assessments indicates that language expectations and criteria are only implicit parts of the assessment although they are an integral part of the overall performance and should be made more explicit.

We also believe that inclusion of a language performance rubric will create stronger alignment with the assessment attributes and performance expectations than the current system. As indicated by the National Science Education Standards (National Research Council, 1996), students are expected to use their content knowledge and the language of science to communicate scientific explanations and ideas effectively. The central problem of content-based assessment is that, although the multidimensionality of the constructs is often represented in the content standards and performance expectations in assessment tasks, these are not reflected in the scoring criteria. Given the multidimensionality of the constructs that the assessments are intended to measure, making the various cognitive demands (e.g., literacy skills) associated with the successful completion of performance assessment tasks more explicit and aligning them with the performance criteria will help increase the validity of the assessment. In addition, acknowledging the various cognitive demands of the assessment and making the performance criteria more explicit will help identify the source of the achievement gap and variability in student performance. As our results indicate, with explicit performance criteria, we were able to differentiate and identify the source of achievement gap between ELs and non-ELs. In genetics at least, ELs achievement gap is attributed to a lack of language proficiency rather than a lack of content knowledge. Under the currently available assessment models, it is difficult to determine whether the EL achievement gap in content areas is attributed to a lack of content knowledge or the language proficiency level.

With more explicit scoring criteria, the assessment can be also used for diagnostic purposes and can provide instructional feedback. Additional performance criteria will increase specificity for evaluating student performance and provide information to guide instruction and learning activities. Clarification of performance expectations and criteria, especially in terms of literacy expectations, will provide teachers with the necessary information to modify and inform instruction. Teachers will be able to use this information to verify the alignment among their learning goals, the standards, and the assessment. Also, with more specificity in student evaluation, teachers will be able to adjust instruction to better meet the individual needs of students. Recent studies have shown that science educators increasingly believe that language, reading, and writing skills play a significant role in understanding science and that the process of teaching literacy skills and science are parallel and mutually supportive of learning (Baker, 1991; McMahon & McCormack, 1998).

Our study also confirms that biology teachers are not engaged in explicit instruction in scientific literacy. Despite the importance of literacy development in the context of content area instruction, teachers are often unprepared and lack resources to integrate literacy instruction into their content area instruction (Hart & Lee, 2003). Proponents of integrated instruction believe that science instruction provides a context for reading and writing for a purpose. Because science learning relies on various cognitive skills, including the ability to access scientific terminology, comprehend scientific texts, and write coherent scientific explanations, science curriculum provides a prime opportunity for teachers to engage students in academic language and literacy development. For ELs, academic language and literacy development is an integral part of content learning. This study contributes to the growing literature advocating for a more integrated approach to teaching science and literacy.

Additionally, previous investigations have shown that there seems to be a mismatch between what teachers believe and report as integrated curriculum and the percentage of time spent on actual integrated literacy activities (Schmidt et al., 1985). Professional development programs providing specific tools and strategies to help teachers facilitate integrated approaches to instruction will be critical for meeting the expected standards and sustaining science development. In the midst of current science education reform, development and investigation of several professional development programs emphasizing integrated literacy instruction in content area instruction (e.g., RA, SEEDS, and SLIA) is currently underway. As the number of ELs continues to rise in our schools, integrated instruction may help facilitate more effective approaches to academic language development. With the current science standards, inadequate support for ELs' language development in content area instruction will be increasingly viewed as an opportunity-to-learn issue as well as an equity issue.

From a research perspective, more rigorous controlled experimental research studies are warranted for determining the impact of integrated instruction on student outcomes. Although there is some anecdotal evidence to support the benefits of integrated instruction on learning, there is limited empirical evidence to support this claim currently. Also, further research is needed to investigate the efficacy of these professional development programs in providing teachers with adequate training.

Finally, given the small sample size, this pilot study had several limitations. First, several of the key instruments used to collect the data in the study are still under development. Although the quality of the instruments is promising based on the reliability and validity evidence, we identified some areas that need further development and refinement for future studies. In teacher assignment ratings, for example, ambiguity in where to look for evidence for quality has created low rater reliability. Although consensus scores were used in the analysis, the low reliability scores require caution when interpreting the results. In addition, the observation protocol used for the validation of the teacher assignment ratings needs further refinement. Without clear alignment between the two measures, it was difficult to provide detailed validity information on the teacher assignment ratings. Also, due to the small number of teachers participating in the study, the generalizability of the results is limited. The small number of sample size also attributed to lack of significant findings.

For the student assessment, the limited amount of content coverage in physiology brings up validity issues regarding the assessment. However, given that physiology is a content area previously covered in the context of other biology classes, we felt compelled to examine this topic area for comparative purposes. Again, for generalizability purposes, further research with standardized and validated instruments is warranted. Based on our pilot study, we have identified the following strategies for improving the quality and the validity of the teacher assignments and ILA:

> Provide specific and detailed instruction to raters on where to find evidence for teacher assignment dimensions. For example, most of the data for teacher assignment ratings are provided by the assignment coversheet. In the descriptions of the dimensions, providing specific item numbers on the coversheet corresponding to each dimension will increase the reliability.
>
> Gather more detailed information about the instructional support provided to students prior to assignment collection through phone interviews and teacher surveys for reliability check on the teacher assignment data.
>
> For cognitive and literacy support dimensions, the ability level of students (determined by the information provided by the teacher on the distribution of performance on the assignment) should be considered in conjunction with the cognitive and literacy challenge to determine the appropriateness of the level of support.
>
> Modify the literacy support dimension to include specific examples of support such as whether the teacher ***models*** and ***provides explicit instruction*** in: (a) conducting focused conversations, (b) metacognitive activities, and (c) reading comprehension strategies, including analyzing text structure and conversation routines to promote student talk and collaborative meaning making activities; and whether the teacher encourages students to assist one another (e.g., explain concepts).
>
> For ILA, additional items to measure reading comprehension should be developed. Administration of only three reading comprehension items was insufficient to determine student's reading comprehension levels.
>
> Standardized content assessment items tapping into basic and prior knowledge related to the targeted topic area should be included as part of the assessment to provide additional information to determine the level of content understanding. Performance on the content knowledge items will provide baseline information about students' level of knowledge and exposure to the key concepts elicited by the assessment prompt.

REFERENCES

Abedi, J. (2004). *Inclusion of students with limited English proficiency in NAEP: Classification and measurement issues* (CRESST Tech. Rep. No. 629). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Abedi, J., Courtney, M., & Leon, S. (2003). *Effectiveness and validity of accommodations for English language learners in large-scale assessment*. (CRESST Tech. Rep. No. 608). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Abedi, J., & Leon, S. (1999). *Impact of students' language background on content-based performance: Analyses of extant data.* Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Abedi, J., Leon, S., & Mirocha, J. (2003). *Impact of students' language background on content-based data: Analyses of extant data* (CRESST Tech. Rep. No. 603). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Aguirre-Muñoz, Z., Boscardin, C., Jones, B., Park, J., Chinen, M., Shin, H., Lee, J., Amabisca, A., & Benner, A. (2005). *Consequences and validity of performance assessment for English language learners: Integrating academic language and ELL instructional needs into opportunity-to-learn measures*. (CRESST Deliverable). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing.* Washington, DC: American Educational Research Association.

Aschbacher, P. R. (1999). *Developing indicators of classroom practice to monitor and support school reform* (CRESST Tech. Rep. No. 513). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

August, D., & Hakuta, K. (Eds.). (1997). *Improving schooling for language-minority children: A research agenda*. Washington, DC: National Academy Press.

Bailey, A. (2000). *OBEMLA Report*. Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Baker, L. (1991). Metacognition, reading, and science education. In C. Santa & D. Alvermann (Eds.), *Science learning: Processes and applications*. Newark, DE: International Reading Association, pp. 2–13.

Baker, E. L. (1994). Making performance assessment work: The road ahead. *Educational Leadership*, *51*(6), 58–62.

Bertrand, C., DiRanna, K., & Janulaw, A. (1999). *Making connections: A guide to implementing science standards.* Sacramento, CA: California Science Teachers Association.

Boulter, C. J., & Gilbert, J. K. (1995). Argument and science education. In P. J. M. Costello & S. Mitchell (Eds.), *Competing and consensual voices: The theory and practice of argumentation.* Clevedon, UK: Multilingual Matters.

Butler, F. A., & Castellon-Wellington, M. (2000). Students' concurrent performance on tests of English language proficiency and academic achievement. In *The validity of administering large-scale content assessments to English language learners: An investigation from three perspectives* (Final Deliverable to OERI/OBEMLA, Contract No. R305B60002; pp. 51–83). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Campbell, N. A., & Reece, J. B. (2005). *Biology* (7th ed.). San Francisco: Pearson/ Benjamin Cummings.

Christie, F. (1986). Writing in schools: Generic structures as ways of meaning. In B. Couture (Ed.), *Functional approaches to writing: Research perspectives* (pp. 221–239). London: Frances Pinter.

Christie, F. (2002). The development of abstraction in adolescence in subject English. In M. Schleppegrell & M. C. Colobi (Eds.), *Developing advanced literacy in first and second language: Meaning with power* (pp. 45–66). London: Routledge.

Chung, G. K., O'Neil, H. F. J., & Herl, H. E. (1999). The use of computer-based collaborative knowledge mapping to measure team processes and team outcomes. *Computers in Human Behavior, 15* (3–4), 463–494.

Davidson, F., & Lynch, B. K. (2002). *Testcraft: A teacher's guide to writing and using language test specifications*. New Haven, CT: Yale University Press.

Duschl, R., & Osborne, J. (2002). Supporting and promoting argumentation discourse. *Studies in Science Education*, *38*, 39–72.

Erduran, S., Simon, S., & Osborne, J. (2004). TAPping into argumentation: Developments in the use of Toulmin's Argument Pattern in studying science discourse. *Science Education*, *88*(6), 915–933.

Fang, Z. (2004). Scientific literacy: A systemic functional linguistics perspective. *Science Education*, *89*(2), 335–347.

Greenberg, J. (Ed.). (2001). *BSCS biology: A molecular approach* (8th ed., Blue Version). Chicago: Everyday Learning.

Halliday, M. A. K. (1994). *An introduction to functional grammar.* (2nd ed.). London: Edward Arnold.

Hart, J., & Lee, O. (2003). Teacher professional development to improve science and literacy achievement of English language learners. *Bilingual Research Journal, 27*(3), 475–501.

Hays, L. J., & Harris, C. D. (2003). Using literacy coaching as a means to change science teachers' attitudes about teaching writing: a case study. *Electronic Journal of Literacy Through Science*, *2*, 1–19.

Martin, K., & Miller, E. (1988). Storytelling and science. *Language Arts*, *65*(3), 255–259.

Matsumura, L. C. (2000). *Using teachers' assignments as an indicator of classroom practice*. (CRESST Tech. Rep. No. 532). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

McCutchen, D. (1986). Domain knowledge and linguistics knowledge in the development of writing ability. *Journal of Memory and Language, 25*(4), 431–444.

McMahon, M., & McCormick, B. B. (1998). To think and act like a scientist: Learning disciplinary knowledge. In C. R. Hynd, S. A. Stahl, & S. M. Glynn, (Eds.), *Learning text across conceptual domains*, pp. 227–262. Mahwah, NJ: Lawrence Earlbaum Associates.

Miller, K. R., & Levine, J. S. (2006). *Prentice Hall biology* (Dragonfly version). NJ and MA: Pearson Prentice Hall.

Morrow, L. M., Pressley, M., Smith, J. K., & Smith, M. (1997). The effect of a literature-based program integrated into literacy and science instruction with children from diverse backgrounds. *Reading Research Quarterly*, *32*(1), 54–76.

National Research Council. (1996). *National Science Education Standards.* Washington DC: National Academy Press.

No Child Left Behind Act of 2001, Pub. L. No. 107–110 § 115 Stat. 1425 (2002).

Pontecorvo, C. (1987). Discussing for reasoning: The role of argument in knowledge construction. In E. De Corte, J. G. L. C. Lodewijks, R. Parmentier, & P. Span (Eds.), *Learning and instruction. A publication of the European Association for Research on Learning and Instruction,* pp. 71–82. Oxford/Leuven, UK: Leuven University Press.

Popham, W. J. (2006). Assessment for learning: An endangered species? *Educational Leadership*, *63*(5), 82–83.

Psathas, G. (Ed.). (1979). *Everyday language: Studies in ethnomethodology.* New York: Irvington.

Ruiz-Primo, M. A., Shavelson, R. J., Hamilton, L., & Klein, S. (2002). On the evaluation of systemic education reform: Searching for instructional sensitivity. *Journal of Research in Science Teaching*, *39*(5), 369–393.

Schleppegrell, M. J. (1998). Grammar as resource: Writing a description. *Research in the Teaching of English, 32*, 182–211.

Schleppegrell, M. J. (2003). *Grammar for writing: Academic language and the ELD standards*. UCLMRI Final Report.

Schleppegrell, M. J. (2004). *The language of schooling: A functional linguistics perspective*. Mahwah, NJ: Lawrence Erlbaum Associates.

Schmidt, W., Roehler, L., Caul, J., Buchman, M., Diamond, B., Solomon, D., et al. (1985). The uses of curriculum integration in languages arts instruction: A study of six classrooms. *Curriculum Studies, 17*(3), 305–320.

Schoenbach, R., Greenleaf, C., Cziko, C., & Hurwitz, L. (1999). *Reading for understanding: A guide to improving reading in middle and high school classrooms*. San Francisco: Jossey-Bass.

Weiss, I. R., Banilower, E. R., McMahon, K. C., & Smith, P. S. (2001). *Report of the 2000 National Survey of Science and Mathematics Education.* Retrieved December, 2006 at http://www.horizonresearch.com/reports/2001/2000survey/full_report.php

APPENDIX A

ILA Test Specifications

Test Specification Biology/Life Sciences – GENETICS 1
Content Standard 5

General Description

It is important for biology students to understand concepts of genetics in order to understand the way life progresses and be able to comprehend a text passage and communicate their thoughts in writing. Within genetics, **students should know that the genetic composition of cells can be altered by the incorporation of exogenous DNA into the cells**. As a basis for understanding this concept students should know the general structures and functions of DNA, RNA, and protein. Additionally, students should know how genetic engineering (biotechnology) is used to produce novel biomedical and agricultural products. Biology students should be able to integrate prior content knowledge together with information from a new text for the purpose of clearly communicating a response to a question relating to the text. Students should be able to draw upon their prior knowledge from other biology content areas.

Students will demonstrate their ability to **read** and **understand** text from appropriate genetics materials by reading a text and answering questions pertaining to reading processes and reading comprehension. Specifically, the target reading comprehension ability entails high-level cognitive skills when engaging with the ideas presented in the text. Students will demonstrate their **knowledge** of genetics by responding to an essay prompt. Specifically, the target content knowledge ability entails connecting the ideas presented in the text to prior content knowledge. Students will also demonstrate their ability to **write** an essay in response to a prompt, using a proper science register to communicate their scientific ideas in an appropriate, authoritative, coherent, and legible manner. In order to demonstrate proficiency in this area, students will:

a. ***Organize*** their piece of writing such that it meets the demands of the genre.
b. ***Present*** their knowledge and ideas using relevant linguistic resources such as:
    i. Correct and appropriate use of scientific terms.
    ii. Use of grammatical items that link scientific ideas and sequences logically and coherently.
    iii. Correct and purposeful use of general nouns for the description of scientific facts or events.
c. ***Consider*** the audience. Students should use grammatical items to maintain an impersonal and objective tone of voice.
d. ***Utilize*** appropriate writing conventions. Students' essays should not contain spelling errors or grammatical mistakes that impede communication.

Prompt Attribute

Text
Students will be asked to read a text.

Requirements for the text:

The text should be scientific and specific to genetics. It may be a text that the students have already seen. Sources include: Scientific American, New Scientist, online scientific articles, and any other scientific materials that contain two or more medium-length paragraphs and pictures or graphs related to genetics.

The text:

Should contain information pertaining to genetic modification and biotechnology.

Should be specific enough that students are required to draw on their prior knowledge of the content domain principles in order to respond to the prompt in a satisfactory manner. In other words, the text should not contain all the information that is being asked for by the prompt.

Should be written in a prose style that is typical of well-written, authentic scientific writing.
Should contain abstract, technical vocabulary as well as various complex linguistic structures in conveying ideas in a logical, precise, and impersonal manner.
May be modified in order to satisfy content-level text requirements (i.e., the text must pertain to general science information that students are tested on yet be specific enough so that it does not explicitly contain all the answers to the questions being asked), but should not be modified in order to provide linguistic support for student reading comprehension per se.

Reading Processes
Students will be asked to answer several open-ended questions about their experience reading the text passage and what strategies they utilized to help them make sense of the text.

Reading Comprehension
Students will be asked to answer several multiple choice questions regarding the content of the text passage.

Requirements for multiple choice questions

This section consists of several multiple choice question items. These items are designed to measure the level of students' reading comprehension for the given text, particularly focusing on the following reading skills

1. Vocabulary Verification
2. Global Paraphrasing
3. Sentence Paraphrasing

Essay in response to a text
Students will be asked to read a text that pertains to genetic modification, which will include a picture or graph and space for student notes. The instructions will indicate that students may make reading comprehension strategy notes in the allotted space. The students will be asked to integrate the information from the text as well as from their prior learning to write an essay.

The item stem will pose a question about the information, statements, viewpoints, or beliefs presented in the text. The stem will ask students to explain their response using the information, statements, viewpoints, or beliefs from the text, as well as their prior knowledge.

Response Attribute

Students will:
1. **Read** the text and examine any accompanying pictures or graphs.
2. **Describe** their experience reading the text passage and what strategies they utilized to help them make sense of the text.
3. **Answer** several multiple choice questions about the content of the text passage.
4. **Write** a coherent and legible essay in **response** to the prompt. The response should include:
    a. Presentation of a thesis with supportive evidence
    b. Interpretations and generalizations to account for how and why things are as they are
    c. Logical organization that presents causes and makes judgments
    d. Emphasis on generalization, classification, and categorization to support judgments
    e. Integration of relevant information from the text, pictures, and graphs, as well as prior biology content knowledge
    f. Use of a proper science register (proper grammar and scientific academic language as opposed to common conversational language) to communicate their ideas in an appropriate, authoritative, and coherent way

Test Specification Biology/Life Sciences – GENETICS 2
Content Standard 4

General Description:

It is important for biology students to understand concepts of genetics in order to understand the way life progresses and be able to comprehend a text passage and communicate their thoughts in writing. Within genetics, **students should know genes are a set of instructions encoded in the DNA sequence of each organism that specify the sequence of amino acids in proteins characteristic of that organism**. As a basis for understanding this concept students should know the general pathway by which proteins are synthesized. Additionally, students should know that specialization of cells in multicellular organisms is usually due to different patterns of gene expression[14]. Biology students should be able to integrate prior content knowledge together with information from a new text for the purpose of clearly communicating a response to a question relating to the text. Specifically, students should be able to draw upon their prior knowledge from other biology content areas.

Students will demonstrate their ability to **read** and **understand** text from appropriate genetics materials by reading a text and answering questions pertaining to reading processes and reading comprehension. Specifically, the target reading comprehension ability entails high-level cognitive skills when engaging with the ideas presented in the text. Students will demonstrate their **knowledge** of genetics by responding to an essay prompt. Specifically, the target content knowledge ability entails connecting the ideas presented in the text to prior content knowledge. Students will also demonstrate their ability to **write** an essay in response to a prompt, using a proper science register to communicate their scientific ideas in an appropriate, authoritative, coherent, and legible manner. In order to demonstrate proficiency in this area, students will:

a. ***Organize*** their piece of writing such that it meets the demands of the genre.
b. ***Present*** their knowledge and ideas using relevant linguistic resources such as:
   i. Correct and appropriate use of scientific terms.
   ii. Use of grammatical items that link scientific ideas and sequences logically and coherently.
   iii. Correct and purposeful use of general nouns for the description of scientific facts or events.
c. ***Consider*** the audience. Students should use grammatical items to maintain an impersonal and objective tone of voice.
d. ***Utilize*** appropriate writing conventions. Students' essays should not contain spelling errors or grammatical mistakes that impede communication.

Prompt Attribute

Text
Students will be asked to read a text.

Requirements for the text:
The text should be scientific and specific to genetics. It may be a text that the students have already seen. Sources include: Scientific American, New Scientist, online scientific articles, and any other scientific materials that contain two or more medium-length paragraphs and pictures or graphs related to genetics.

The text:

Should contain information pertaining to genetics and DNA testing.
Should be specific enough that students are required to draw on their prior knowledge of the content domain principals in order to respond to the prompt in a satisfactory manner.

[14] CDE B/LS Standard 4 (a, d) www.cde.ca.gov/be/st/ss/scbiology

In other words, the text should not contain all the information that is being asked for by the prompt.

Should be written in a prose style that is typical of well-written, authentic scientific writing.

Should contain abstract, technical vocabulary as well as various complex linguistic structures in conveying ideas in a logical, precise, and impersonal manner.

May be modified in order to satisfy content-level text requirements (i.e., the text must pertain to general science information that students are tested on yet be specific enough so that it does not explicitly contain all the answers to the questions being asked), but should not be modified in order to provide linguistic support for student reading comprehension per se.

Reading Processes
Students will be asked to answer several open-ended questions about their experience reading the text passage and what strategies they utilized to help them make sense of the text.

Reading Comprehension
Students will be asked to answer several multiple choice questions regarding the content of the text passage.

Requirements for multiple choice questions

This section consists of several multiple choice question items. These items are designed to measure the level of students' reading comprehension for the given text, particularly focusing on the following reading skills

1. Vocabulary Verification
2. Global Paraphrasing
3. Sentence Paraphrasing

Essay in response to a text
Students will be asked to read a text that pertains to DNA testing, which will include a picture or graph and space for student notes. The instructions will indicate that students may make reading comprehension strategy notes in the allotted space. The students will be asked to integrate the information from the text as well as from their prior learning to write an explanation and/or exposition.

The item stem will pose a question about the information, statements, viewpoints, or beliefs presented in the text. The stem will ask students to explain their response using the information, statements, viewpoints, or beliefs from the text, as well as their prior knowledge.

Response Attribute

Students will:

1. **Read** the text and examine any accompanying pictures or graphs.
2. **Describe** their experience reading the text passage and what strategies they utilized to help them make sense of the text.
3. **Answer** several multiple choice questions about the content of the text passage.
4. **Write** a coherent and legible essay in **response** to the prompt. The response should include:
    a. Presentation of a thesis with supportive evidence
    b. Interpretations and generalizations to account for how and why things are as they are
    c. Logical organization that presents causes and makes judgments
    d. Emphasis on generalization, classification, and categorization to support judgments
    e. Integration of relevant information from the text, pictures, and graphs, as well as prior biology content knowledge
    f. Use of a proper science register (proper grammar and scientific academic language as opposed to common conversational language) to communicate their ideas in an appropriate, authoritative, and coherent way.

Test Specification Biology/Life Sciences – PHYSIOLOGY 1
Content Standard 10

General Description:

It is important for biology students to understand concepts of physiology in order to understand how the human body combats illness and disease. Within physiology, **students should know that the body has a variety of mechanisms to interfere with and destroy invading pathogens (bacteria, parasites, and viruses) that cause human illnesses and disease.** As a basis for understanding this concept students should know the role of the skin in providing nonspecific defense. Students should also know the role of antibodies in the body's response to infection. Biology students should be able to integrate prior content knowledge together with information from a new text for the purpose of clearly communicating a response to a question relating to the text. Students should be able to draw upon their prior knowledge from other biology content areas.

Students will demonstrate their ability to **read** and **understand** text from appropriate immune system physiology materials by reading a text and answering questions pertaining to reading processes and reading comprehension. Specifically, the target reading comprehension ability entails high-level cognitive skills when engaging with the ideas presented in the text. Students will demonstrate their **knowledge** of immune system physiology by responding to an essay prompt. Specifically, the target content knowledge ability entails connecting the ideas presented in the text to prior content knowledge. Students will also demonstrate their ability to **write** an essay in response to a prompt, using a proper science register to communicate their scientific ideas in an appropriate, authoritative, coherent, and legible manner. In order to demonstrate proficiency in this area, students will:

a. ***Organize*** their piece of writing such that it meets the demands of the genre.
b. ***Present*** their knowledge and ideas using relevant linguistic resources such as:
    i. Correct and appropriate use of scientific terms.
    ii. Use of grammatical items that link scientific ideas and sequences logically and coherently.
    iii. Correct and purposeful use of general nouns for the description of scientific facts or events.
c. ***Consider*** the audience. Students should use grammatical items to maintain an impersonal and objective tone of voice.
d. ***Utilize*** appropriate writing conventions. Students' essays should not contain spelling errors or grammatical mistakes that impede communication.

Prompt Attribute

Text
Students will be asked to read a text.

Requirements for the text:

The text should be scientific and specific to immune system physiology. It may be a text that the students have already seen. Sources include: Scientific American, New Scientist, online scientific articles, and any other scientific materials that contain two or more medium-length paragraphs and pictures or graphs related to immune system physiology.

The text:

Should contain information pertaining to the immune system and physiology.
Should be specific enough that students are required to draw on their prior knowledge of the content domain principles in order to respond to the prompt in a satisfactory manner. In other words, the text should not contain all the information that is being asked for by the prompt.

Should be written in a prose style that is typical of well-written, authentic scientific writing.
Should contain abstract, technical vocabulary as well as various complex linguistic structures in conveying ideas in a logical, precise, and impersonal manner.
May be modified in order to satisfy content-level text requirements (i.e., the text must pertain to general science information that students are tested on yet be specific enough so that it does not explicitly contain all the answers to the questions being asked), but should not be modified in order to provide linguistic support for student reading comprehension per se.

Reading Processes
Students will be asked to answer several open-ended questions about their experience reading the text passage and what strategies they utilized to help them make sense of the text.

Reading Comprehension
Students will be asked to answer several multiple choice questions regarding the content of the text passage.

Requirements for multiple choice questions

This section consists of several multiple choice question items. These items are designed to measure the level of students' reading comprehension for the given text, particularly focusing on the following reading skills:

1. Vocabulary Verification
2. Global Paraphrasing
3. Sentence Paraphrasing

Essay in response to a text
Students will be asked to read a text that pertains to the immune system physiology, which will include a picture or graph and space for student notes. The instructions will indicate that students may make reading comprehension strategy notes in the allotted space. The students will be asked to integrate the information from the text as well as from their prior learning to write an essay.

The item stem will pose a question about the information, statements, viewpoints, or beliefs presented in the text. The stem will ask students to explain their response using the information, statements, viewpoints, or beliefs from the text, as well as their prior knowledge.

Response Attribute

Students will:
1. **Read** the text and examine any accompanying pictures or graphs.
2. **Describe** their experience reading the text passage and what strategies they utilized to help them make sense of the text.
3. **Answer** several multiple choice questions about the content of the text passage.
4. **Write** a coherent and legible essay in **response** to the prompt. The response should include:
    a. Presentation of a thesis with supportive evidence
    b. Interpretations and generalizations to account for how and why things are as they are
    c. Logical organization that presents causes and makes judgments
    d. Emphasis on generalization, classification, and categorization to support judgments
    e. Integration of relevant information from the text, pictures, and graphs, as well as prior biology content knowledge.
    f. Use of a proper science register (proper grammar and scientific academic language as opposed to common conversational language) to communicate their ideas in an appropriate, authoritative, and coherent way.

Test Specification Biology/Life Sciences – PHYSIOLOGY 2
Content Standard 10

General Description

It is important for biology students to understand concepts of physiology in order to understand how the human body combats illness and disease. Within physiology, **students should know that as a result of the coordinated structures and functions of organ systems, the internal environment of the human body remains relatively stable (homeostatic) despite changes in the outside environment**. As a basis for understanding this concept students should know how the nervous system mediates communication between different parts of the body and the body's interactions with the environment. Additionally, students should also know how feedback loops in the nervous and endocrine systems regulate conditions in the body. Biology students should be able to integrate prior content knowledge together with information from a text for the purpose of clearly communicating a response to a question relating to the text. Students should be able to draw upon their prior knowledge from other biology content areas.

Students will demonstrate their ability to **read** and **understand** text from appropriate genetics materials by reading a text and answering questions pertaining to reading processes and reading comprehension. Specifically, the target reading comprehension ability entails high-level cognitive skills when engaging with the ideas presented in the text. Students will demonstrate their **knowledge** of genetics by responding to an essay prompt. Specifically, the target content knowledge ability entails connecting the ideas presented in the text to prior content knowledge. Students will also demonstrate their ability to **write** an essay in response to a prompt, using a proper science register to communicate their scientific ideas in an appropriate, authoritative, coherent, and legible manner. In order to demonstrate proficiency in this area, students will:

a. ***Organize*** their piece of writing such that it meets the demands of the genre.
b. ***Present*** their knowledge and ideas using relevant linguistic resources such as:
    i. Correct and appropriate use of scientific terms.
    ii. Use of grammatical items that link scientific ideas and sequences logically and coherently.
    iii. Correct and purposeful use of general nouns for the description of scientific facts or events.
c. ***Consider*** the audience. Students should use grammatical items to maintain an impersonal and objective tone of voice.
d. ***Utilize*** appropriate writing conventions. Students' essays should not contain spelling errors or grammatical mistakes that impede communication.

Prompt Attribute

Text
Students will be asked to read a text.

Requirements for the text:
The text should be scientific and specific to genetics. It may be a text that the students have already seen. Sources include: Scientific American, New Scientist, online scientific articles, and any other scientific materials that contain two or more medium-length paragraphs and pictures or graphs related to genetics.

The text:

Should contain information pertaining to genetic modification and biotechnology.
Should be specific enough that students are required to draw on their prior knowledge of the content domain principles in order to respond to the prompt in a satisfactory manner. In other words, the text should not contain all the information that is being asked for by the prompt.

Should be written in a prose style that is typical of well-written, authentic scientific writing.
Should contain abstract, technical vocabulary as well as various complex linguistic structures in conveying ideas in a logical, precise, and impersonal manner.
May be modified in order to satisfy content-level text requirements (i.e., the text must pertain to general science information that students are tested on yet be specific enough so that it does not explicitly contain all the answers to the questions being asked), but should not be modified in order to provide linguistic support for student reading comprehension per se.

Reading Processes
Students will be asked to answer several open-ended questions about their experience reading the text passage and what strategies they utilized to help them make sense of the text.

Reading Comprehension
Students will be asked to answer several multiple choice questions regarding the content of the text passage.

Requirements for multiple choice questions

This section consists of several multiple choice question items. These items are designed to measure the level of students' reading comprehension for the given text, particularly focusing on the following reading skills

1. Vocabulary Verification
2. Global Paraphrasing
3. Sentence Paraphrasing

Essay in response to a text
Students will be asked to read a text that pertains to genetic modification, which will include a picture or graph and space for student notes. The instructions will indicate that students may make reading comprehension strategy notes in the allotted space. The students will be asked to integrate the information from the text as well as from their prior learning to write an essay.

The item stem will pose a question about the information, statements, viewpoints, or beliefs presented in the text. The stem will ask students to explain their response using the information, statements, viewpoints, or beliefs from the text, as well as their prior knowledge.

Response Attribute

Students will:

1. **Read** the text and examine any accompanying pictures or graphs.
2. **Describe** their experience reading the text passage and what strategies they utilized to help them make sense of the text.
3. **Answer** several multiple choice questions about the content of the text passage.
4. **Write** a coherent and legible essay in **response** to the prompt. The response should include:
    a. Presentations of a thesis with supportive evidence
    b. Interpretations and generalizations to account for how and why things are as they are
    c. Logical organization that presents causes and makes judgments
    d. Emphasis on generalization, classification, and categorization to support judgments
    e. Integration of relevant information from the text, pictures, and graphs, as well as prior biology content knowledge.
    f. Use of a proper science register (proper grammar and scientific academic language as opposed to common conversational language) to communicate their ideas in an appropriate, authoritative, and coherent way.

APPENDIX B

Text Analysis of Biology Textbooks

Text Analysis of Biology Textbooks

1. High school biology texts frequently include technical vocabulary (e.g., *eukaryotic gene regulation*, *DNA fingerprint*, etc.).

2. Various verb types are used in explaining biological concepts, objects, and processes and also in stating historical findings and events in the field of biology. These verb types include actions verbs (e.g., *study*, *produce*, *cross*, *undergo*, etc.), mental verbs (e.g., *wanted to study*, *assumed*, etc.), and relational verbs (e.g., *be*). Particularly, relational verbs frequently occur in defining concepts and technical vocabulary (e.g., *A trait is X*, *X is called Y*, etc.). Except for historical remarks (e.g., *Mendel studied X*), scientific facts are expressed in present tense.

3. Passive sentence structures are frequently used, thereby minimizing any indication of human involvement in the working of the natural world and emphasizing instead non-human subjects such as *gene, cell, element,* etc. Some of these non-human subjects appear as abstract noun forms such as nominalizations (expressing a verb as a noun, such as forming the noun *destruction* from the verb *destroy*).

4. In relation to 3, expanded noun phrases are frequently found in subject position, creating a logical thematic progression (e.g., *common patterns of genetic control*, *the striking similarity of genes that control development*, etc.)

5. Various subordinate clauses (e.g., *when a copy of the mouse gene was inserted*), prepositional phrases, and participial phrases are positioned before and/or after the main clause. These are used to mark logical sequences within a process that is explained. This type of clause-linking strategy makes scientific explanations concise and straightforward.

6. In relation to 5, the use of easy connectors (e.g., *and*, *so*, *but*) is limited.

7. Sentences are in declarative form because textbooks mainly involve factual statements. However, interrogative sentences and imperatives do appear in order to solicit readers' active participation (e.g., *think for a moment about X*, *remember X*, *why must X do Y?*). Similarly, first- and second-person reference forms occasionally occur as a device for drawing readers' attention. Overall, however, first- and second-person references rarely occur.

APPENDIX C

Text Evaluation Guide for Teachers

Dear Biology Teacher,

Thank you for agreeing to take some time to help us evaluate the student assessment prompts (including the text passages, directions, etc.). As you do this, we would like you to consider YOUR typical/average biology student when rating the following dimensions. The dimensions are included to the left of the question to serve as a guide to your evaluation. Please make any additional comments that you have in the area below each question. We really appreciate the feedback! Enjoy your gift certificate! Thank you! The CRESST Biology Team

Ratings Scale:	**5 Confusing**		**3 Clear**		**1 Overly explanatory**
1. The assessment directions are:	5	4	3	2	1

Tells us if it is clear what the students are asked to do.	Comments:

	5 Too hard		**3 Just right**		**1 Too easy**
2. The language used in the assessment directions is:	5	4	3	2	1

Tells us if you think the language is too challenging	Comments:

	5 Too hard		**3 Just right**		**1 Too easy**
3. The language used in the article is:	5	4	3	2	1

Tells us if you think the language being used is at grade level for your students	Comments:

	5 Too hard		**3 Just right**		**1 Too easy**
4. The content discussed in the article is:	5	4	3	2	1

Tells us if the information in the text is at grade level for your students	Comments:

	5 Too hard		**3 Just right**		**1 Too easy**
5. Your overall rating of the student assessment:	5	4	3	2	1

Tells us if you think the question is above or below grade level for your students	Comments:

APPENDIX D

ILA Prompts

High School Biology Performance Assessment

Genetics 1

Student	
Teacher	
School	
Period	

CRESST

National Center for Research on Evaluation, Standards, and Student Testing

Genetics

Part A: Reading - Genetic Modification

In this part of the test, you will read a passage on traditional cross-breeding and genetic engineering. As you read the passage, you should think about what you have learned about genetics. You may underline, take notes, and write in the margin to help you make sense of what you are reading. When you have finished, continue to part B of the booklet.

Traditional Breeding vs. Genetic Engineering[15]

History of Cross-breeding
The link between biotechnology and food dates back over 3,000 years. When biotechnology began in 1800 B.C., yeast was used to leaven bread and ferment wine. Deliberate cross-breeding, another technique of biotechnology, began only a few hundred years ago in the 1860's (by Gregor Mendel). Foods such as potatoes, corn, tomatoes, wheat, oat, and rice are all products of traditional cross-breeding.

History of Genetic Engineering
The most recent technique in biotechnology was developed in 1973 and is called genetic engineering. This refers to the ability to transfer genetic information between plants using molecular technology.

In recent times, crossbreeding has led to a vast increase in the variety and quality of foods available to consumers.

In genetic engineering, one or more genes are removed from one organism and added to the genome of another organism. A gene holds information that will give the organism a trait. Genetic engineering is one type of genetic modification. Traditional plant cross-breeding also modifies the genetic composition of plants. Every time people cross two plants in order to improve their traits, they are genetically modifying the plants.

Limitations of Cross-breeding
Plant cross-breeding is an important tool, but it has limitations. First, cross-breeding can only be done between two plants that can sexually mate with each other. This limits the new traits that can be added to those that already exist in that species. Second, when plants are mated (crossed), many unwanted traits are also transferred along with the trait of interest. All 100,000 or so of each plant's genes are mixed. Since traditional plant breeders only want a few genes transferred, they usually spend 10 to 12 years backcrossing the new plants with the original plants to keep the desired traits while breeding out the tens of thousands of undesired traits.

---

[15] http://citnews.unl.edu/nutrition/html/lesson.shtml?lesson_id=991751218 (Albrecht, J. and Flak, J.) and http://www.quackwatch.org/03HealthPromotion/gmo.html

Advantages of Genetic engineering

Genetic engineering is not bound by the limitations of traditional cross-breeding. Genetic engineering physically removes the DNA from one organism and transfers the gene(s) for one or more traits into another organism. Crossing is not necessary, so there is no longer a need for plants to 'sexually' reproduce. Therefore, traits from any living organism can be transferred into a plant. This method is more specific because a single trait can be added to a plant, eliminating the need to backcross all the undesirable traits.

Part B: Reading Process

In this part of the test, you will answer some questions about your experience reading the passage. Write your answers in the spaces provided below each question. When you have finished, continue to part C of the booklet.

1. **In your own words, write a short summary (one to two sentences) of this passage.**

2. **What thoughts and ideas came to mind as you read this?**

3. **What did you do to help you understand the reading?**

4. **What questions or problems do you still have with this passage?**

5. **How well would you say you understood this passage?**

Part C: Reading Comprehension

In this part of the test, you will answer some questions about the content of the passage. Choose what you think is the best answer for each question, and circle the letter next to the answer you choose. You may choose only one answer for each question. Remember, you can look back at the passage in Part A. When you have finished, continue to Part D of the booklet.

1. According to the passage, genetic engineering:

   A. Is more precise than cross-breeding but takes much longer to complete.

   B. Was developed by Gregor Mendel.

   C. Involves the removal and transfer of DNA from one organism into another.

   D. Manufactures DNA to create new organisms.

2. Which statement below most closely expresses the main idea of the passage?

   A. Scientists' ability to improve the food people eat has progressed through the years to the point that with genetic engineering, scientists can alter food quickly and efficiently.

   B. Traditional cross-breeding was a scientific breakthrough in the late 1800s, but because it was so time-consuming, cross-breeding was not used to alter any of the common foods people eat today.

   C. Scientists prefer traditional cross-breeding because it allows them to transfer large quantities of genetic information from one organism to another, giving them a wide selection of genes from which to choose.

   D. Genetic engineering used for the purpose of creating new organisms is currently a moral and ethical issue.

3. According to the passage, in traditional cross-breeding scientists:

   A. Transfer only a few genes at a time.

   B. Are not able to cross-breed plants that reproduce sexually.

   C. Require many years to develop a new plant with the desired traits.

   D. Attempt to transfer out unwanted traits and transfer in desired traits from the parent generation to the offspring generation.

Part D: Writing

In this part of the test, you will write an essay about genetic modification. Write your essay on the following lined pages. Write as neatly as you can. If you want to change a word, cross it out and write a new word above it. When you have finished, you may close the booklet.

Essay

Imagine that your school is developing a new high school textbook to use next year and that it will be written by students. You have been asked to write an essay to be included in the textbook. The editor of the textbook has asked you to include information in your essay that you have learned during the year in biology class as well as information from the "Traditional Breeding vs. Genetic Engineering" passage in Part A. Specifically, your assignment is to explain how a scientist, using genetic engineering, would alter long grain white rice to make it less appetizing to insects that normally like to eat it. In your essay, include information on:

A. Why inserting the new or different DNA into a cell alters the genetic composition of that cell.
B. How this process of genetic modification is completed.
C. At what point in time a new gene should be inserted into the genome of an organism in order for the new trait to be expressed.

In your essay, make sure to:

1. **Explain the most important aspects** of genetic engineering used to alter an organism.
2. **Organize your ideas logically** so that someone reading your essay can understand how genetic engineering can be used to change an organism.
3. **Incorporate information** that you learned in class and from the passage in Part A to support these important ideas.
4. **Make connections** between genetic engineering and an organism's expression of different traits.
5. **Include** an introduction, body, and conclusion in your essay, and write at least five paragraphs.

After you have finished writing, you may want to re-read your essay and make corrections.

High School Biology Performance Assessment

Genetics 2

Student	
Teacher	
School	
Period	

CRESST
National Center for Research on Evaluation, Standards, and Student Testing

Genetics

Part A: Reading - Genetic Conservation

In this part of the test, you will read a passage on animal conservation. As you read the passage, you should think about what you have learned about genetics. You may underline, take notes, and write in the margin to help you make sense of what you are reading. When you have finished, continue to part B of the booklet.

Great Whites on the Menu[16]

Hunting Great White Sharks

For years, great white sharks have been trophy-hunted for their large jaws and teeth. Now, thanks to DNA testing, conservationists have evidence that even the smaller members of the species are being killed, not for their jaws and teeth, but rather for their fins, which are probably sold for food in Asia. This evidence could be useful in winning greater international protection for the species.

Controversy over Protecting Great Whites

A few countries, including the United States, have outlawed the killing of great white sharks to help prevent the extinction of the species. In 2004, in order to protect great white sharks, the buying and selling of products (such as fins) from great whites began to be monitored and regulated. Some countries, however, have argued that stronger protection was unnecessary, because they believed the sharks were not being frequently killed to be sold for profit. Unfortunately, reliable information to settle the question was lacking until now.

DNA Evidence

Now, there is conclusive evidence that small great white sharks are regularly being killed. Four years ago, researchers developed a DNA test that could help identify various shark species. Not long afterward, they were contacted by law enforcement agents, who had confiscated 900 kilograms of dried shark fins from a ship headed for Asia from the United States.

Taking stock.
Genetic tests are helping reveal which sharks are killed for their fins.

The group reported that all 21 fins in the bag had DNA fingerprints of great white sharks. By analyzing the shape and size of the fins, the researchers concluded that 18 out of the 21 were from small sharks. Since these fins would make puny trophies, they could be intended instead for the kitchen. (Shark fin soup is a popular dish in some countries.) The fact that a large U.S. exporter took the risk of illegally exporting them suggested to the researchers that there is a valuable market even for small fins from great white sharks.

---

16 http://sciencenow.sciencemag.org/cgi/content/full/2006/113/2

According to one researcher, the genetic tests are extremely valuable for animal conservation and are the only way to save the great white shark species. In order to enforce protection laws, the researcher explained, scientists have to be able to identify the different types of shark species from a dried-up fin.

Part B: Reading Process

In this part of the test, you will answer some questions about your experience reading the passage. Write your answers in the spaces provided below each question. When you have finished, continue to part C of the booklet.

1. **In your own words, write a short summary (one to two sentences) of this passage.**

2. **What thoughts and ideas came to mind as you read this?**

3. **What did you do to help you understand the reading?**

4. **What questions or problems do you still have with this passage?**

5. **How well would you say you understood this passage?**

Part C: Reading Comprehension

In this part of the test, you will answer some questions about the content of the passage. Choose what you think is the best answer for each question, and circle the letter next to the answer you choose. You may choose only one answer for each question. Remember, you can look back at the passage in Part A. When you have finished, continue to part D of the booklet.

1. Conservationists are people who:

   A. Hunt animals to earn a living.

   B. Try to protect animals and the environment.

   C. Are working to discover ways to reduce the number of plant and animal species on the planet.

   D. Protect the rights of tribal hunters who rely on hunting for a living.

2. Which statement below most closely expresses the main idea of the passage?

   A. DNA tests are being used to identify endangered species like the great white shark.

   B. The controversy over whether the small great white sharks are being frequently killed for commercial purposes has been finally resolved by evidence provided by the DNA tests.

   C. Only a few countries have currently outlawed killing small great white sharks.

   D. Small great white sharks are often killed for their jaws and teeth.

3. According to the passage, DNA tests are used to:

   A. Identify various shark species.

   B. Detect the mercury level in sharks.

   C. Link small great white sharks to their predators.

   D. Identify the various commercial uses of small great white sharks.

Part D: Writing

In this part of the test, you will write an essay about genetic conservationism. Write your essay on the following lined pages. Write as neatly as you can. If you want to change a word, cross it out and write a new word above it. When you have finished, you may close the booklet.

Essay

Imagine that your school is developing a new high school textbook to use next year and that it will be written by students. You have been asked to write an essay to be included in the textbook. The editor of the textbook has asked you to include information in your essay that you have learned during the year in biology class as well as information from the "Great Whites on the Menu" passage in Part A. Specifically, your assignment is to explain how a conservation geneticist can identify a great white shark fin by examining its DNA fingerprint. In your essay, include information on:

A. What is meant by the term "DNA fingerprint".
B. How would conservation geneticists be able to use a DNA test to determine the identity of an organism?
C. What makes the DNA of one organism different from the DNA of other types of organisms?

In your essay, make sure to:

1. **Explain the most important aspects** of genetic conservationism and DNA testing.
2. **Organize your ideas logically** so that someone reading your essay can understand how DNA "fingerprints" can be used to determine the identity of an organism.
3. **Incorporate information** that you learned in class and from the passage in Part A to support these important ideas.
4. **Make connections** between genetic conservation and DNA testing.
5. **Include** an introduction, body, and conclusion in your essay, and write at least five paragraphs.

After you have finished writing, you may want to re-read your essay and make corrections.

High School Biology Performance Assessment

Physiology 1

Student	
Teacher	
School	
Period	

CRESST

National Center for Research on Evaluation, Standards, and Student Testing

Physiology

Part A: Reading - Allergies

In this part of the test, you will read a passage on allergies and the immune system. As you read the passage, you should think about what you have learned about physiology. You may underline, take notes, and write in the empty spaces to help you make sense of what you are reading. When you have finished, continue to part B of the booklet.

What Are Allergies?[17]

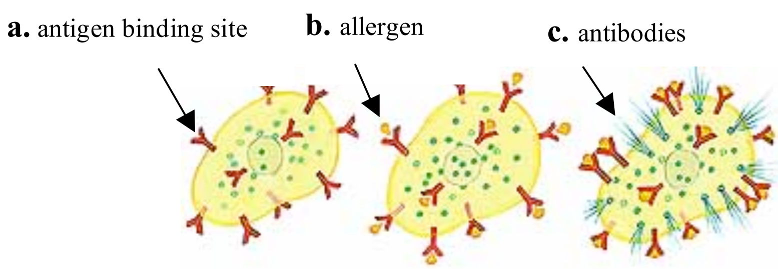
Rarely is a medical term used as frequently and casually as "allergy." It pops up in everyday conversation to describe what people don't like — school, work, or just about anything else.

Medical definition of "allergy"

Medically speaking, "allergy" has a very specific meaning: it is a biologic reaction to something a person's body does not like: allergens. Allergies are an abnormal sensitivity to allergens, which can be ingested while eating, inhaled while breathing, or absorbed through the skin by touch. Most people can tolerate them without trouble. Usually, a person's immune system does not waste its time by reacting against nontoxic substances in the environment. Instead, it uses its various weapons to fight the viruses, bacteria, fungi, and other parasites that threaten the body's health.

Symptoms of allergies

In people with allergies, the set of weapons that normally protects against harmful worms and parasites attacks perfectly harmless substances such as ragweed pollen from grass, animal dander, or certain foods.



***a.** White blood cells in the immune response display antigen binding sites that fit the shape of a specific allergen.*

***b.** Allergens bind to the antigen-binding site on the white blood cells.*

***c.** White blood cells producing antibodies in response to the allergen.*

Steps in the development of an allergic reaction

An allergic reaction is what happens when a person's immune system tries to defend itself against what it believes is a hostile invader. During a period of *sensitization*, the time when a person is exposed to a specific allergen, white blood cells that come in contact with an allergen produce proteins called *antibodies*. If a person were fighting an invasion by a parasite, this would be the first step in a series of events that could lead to the parasite's death. In the case of a brief exposure to an allergen, only a few antibodies are produced and

[17] www.intelihealth.com/IH/ihtIH/WSIHW000/7945/8218/174513.html?d=dmtContent

these may not trigger a full reaction at first. In either case, the immune system is ready to quickly produce a large number of antibodies the next time the same allergen comes along. These antibodies can cause sneezing, runny nose, itching, and other symptoms of an allergic reaction. If a person experiences a more intense reaction with congestion of the tissue, symptoms might include nasal congestion (a stuffy nose), persisting asthma symptoms, or skin rashes.

Part B: Reading Process

In this part of the test, you will answer some questions about your experience reading the passage. Write your answers in the spaces provided below each question. When you have finished, continue to part C of the booklet.

1. **In your own words, write a short summary (one to two sentences) of this passage.**

2. **What thoughts and ideas came to mind as you read this?**

3. **What did you do to help you understand the reading?**

4. **What questions or problems do you still have with this passage?**

5. **How well would you say you understood this passage?**

Part C: Reading Comprehension

In this part of the test, you will answer some questions about the content of the passage. Choose what you think is the best answer for each question, and circle the letter next to the answer you choose. You may choose only one answer for each question. Remember, you can look back at the passage in Part A. When you have finished, continue to part D of the booklet.

1. According to the passage, people who have allergies:

   A. Are frequently allergic to worms, parasites, and fungi.

   B. Have weak immune systems.

   C. Have immune systems that react to harmless substances.

   D. Are able to pass on their allergies to other people through sneezing and a runny nose.

2. Choose the statement below that best characterizes the main idea of the passage.

   A. Most people with allergies react strongly the first time they are exposed to an allergen and so they need to be very careful about what substances they come in contact with.

   B. Allergies occur when the body's defense system reacts to harmless substances after frequent exposure. Because of this, a person usually needs to be exposed to an allergen several times before his/her immune system reacts strongly

   C. Allergies are a product of the body's immune system and affect most people, especially those frequently exposed to worms, parasites, and fungi.

   D. It is best to avoid people with allergies because they are contagious and can cause allergies in someone who didn't previously have them.

3. According to the passage, which of the following statements is true of allergies?

   A. "Allergy" is not a medical term.

   B. Antibodies are produced before exposure to an allergen.

   C. Antibodies always trigger a full reaction as a result of exposure to any allergen.

   D. The period of sensitization begins when a person first comes in contact with an allergen.

Part D: Writing

In this part of the test, you will write an essay about allergies and the immune system. Write your essay on the following lined pages. Write as neatly as you can. If you want to change a word, cross it out and write a new word above it. When you have finished, you may close the booklet.

Essay

Imagine that your school is developing a new high school textbook to use next year and that it will be written by students. You have been asked to write an essay to be included in the textbook. The editor of the textbook has asked you to include information in your essay that you have learned during the year in biology class as well as information from the "What are allergies?" passage in Part A. Specifically, your assignment is to explain:

A. Why the immune system of a person with an allergy to cat dander launches an attack against the allergen.

B. The role of the person's skin in the immune response against cat dander.

C. How the immune system would respond the first time the person is exposed to cat dander, and how the immune system would respond with future exposure to cat dander.

In your essay, make sure to:

1. **Explain, in your own words, the most important aspects** of the immune system and an allergy attack.

2. **Organize your ideas logically** so that a person reading your essay can understand why and how the immune system may launch an attack against an allergen.

3. **Incorporate information** that you learned in class and from the passage in Part A to support these important ideas.

4. **Make connections** between the immune system, allergens, and allergic reactions.

5. **Include** an introduction, body, and conclusion in your essay, and write at least five paragraphs.

After you have finished writing, you may want to re-read your essay and make corrections.

High School Biology Performance Assessment

Physiology 2

Student	
Teacher	
School	
Period	

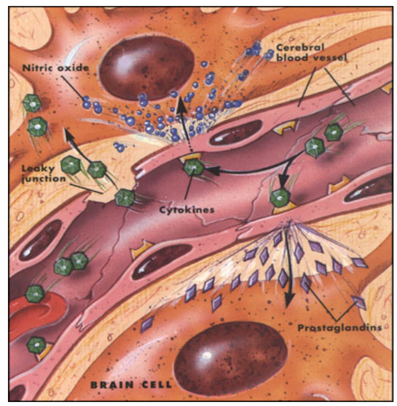
CRESST
National Center for Research on Evaluation, Standards, and Student Testing

Part A: Reading – Immune System

In this part of the test, you will read a passage on links between the mind and body. As you read the passage, you should think about what you have learned about physiology. You may underline, take notes, and write in the margin to help you make sense of what you are reading. When you have finished, continue to part B of the booklet.

The Mind-Body Link[18]

Sometimes it seems like everything is against you. You slip on the ice. Your dog bites you. Then, only a week before finals, you catch your sister's cold. A fever and the "blahs" compound your feelings of bad luck. These symptoms, however, are a sign that something is on your side: your immune system. Your brain is on your side, too, according to an increasing number of studies.



One way immune molecules "talk" to the brain is through the blood. The large molecules are too big to cross from the blood to the brain, but they may be able to slip across tiny holes in the walls of blood vessels. Another way they get their message across is by attaching to special areas on blood vessels and triggering the production of molecules that can di-rectly relay messages to brain cells.

The Immune System

The immune system battles countless enemies. Its wrath is unleashed on viruses, bacteria, parasites, and other foreign molecules that make it past body borders and try to stake a claim. The immune system also combats abnormalities that arise inside the body, such as cancer cells.

Researchers once believed that the immune system was an entirely independent entity in the body. Now an increasing number of studies show that the immune system is tightly connected to the nervous system (which is related to nerve and brain function), as well as to the endocrine system, which is related to hormones. It appears that their three-way communication is necessary for an adequate defense of the body and brain.

Connections among the Immune, Nervous, & Endocrine Systems

Starting in the 1980s, researchers found evidence of strong connections between the immune, nervous, and endocrine systems. First they identified direct links between nerve fibers and immune organs. More recently, they determined that hormones of the endocrine system help the immune and nervous systems defend the body. For example, stress hormones can initiate actions in the brain and immune system in response to injury or germs. This stress response acts as an immune system regulator. It can restrict the immune system response so it does not go overboard. Scientists also recently discovered that immune molecules known as *cytokines* can initiate brain actions. For example, some cytokines help the body heal by sending messages to the brain that set off a series of sickness responses, such as fever. The high body temperature of a fever is thought to create an unfavorable environment for the foreign invaders. The immune molecules also

[18] www.sfn.org

can trigger feelings of sluggishness, fatigue, and loss of appetite. The behaviors can keep sick people out of harm's way until they feel better. Researchers found that cytokines can activate certain nerves for quick brain activation or set off actions from the blood (see illustration).

Implications for understanding these connections

The increasing number of links that researchers are discovering between the immune, nervous, and endocrine systems is leading them to investigate whether excess stress or too little stress can abnormally alter the immune defenses. Others are examining how defects in this intricate system can possibly lead to autoimmune disorders, in which the immune system attacks the body. In addition, scientists are continuing to map the cross-communication network to identify new ways to improve diagnosis and head off disease.

Part B: Reading Process

In this part of the test, you will answer some questions about your experience reading the passage. Write your answers in the spaces provided below each question. When you have finished, continue to part C of the booklet.

1. **In your own words, write a short summary (one to two sentences) of this passage.**

2. **What thoughts and ideas came to mind as you read this?**

3. **What did you do to help you understand the reading?**

4. **What questions or problems do you still have with this passage?**

5. **How well would you say you understood this passage?**

Part C: Reading Comprehension

In this part of the test, you will answer some questions about the content of the passage. Choose what you think is the best answer for each question, and circle the letter next to the answer you choose. You may choose only one answer for each question. Remember, you can look back at the passage in Part A. When you have finished, continue to part D of the booklet.

1. Choose the statement below that best characterizes the main idea in the passage.

   A. The immune, endocrine, and nervous systems all work together to protect the body against illness and injury.

   B. The immune system is now believed to be the body's prime defense mechanism and does not work closely with either the endocrine system or the nervous system to protect the body against illness and injury, as previously thought.

   C. The immune system, once believed to protect the body, is now thought to be a system that works to encourage disease and produce abnormal cells, compromising a person's ability to remain healthy.

   D. Not enough information is known about the immune, nervous, and endocrine systems to determine their roles in defending against stress.

2. According to the passage, defects in a person's immune system can result in:

   A. A breakdown in communication between the nervous system and the immune system.

   B. The immune system developing the ability to fight off disease and injury.

   C. The release of hormones by the endocrine system into the body to support the immune system.

   D. The taking over of the functions of the nervous and endocrine systems by the immune system.

3. According to the passage, cytokines are:

   A. Immune molecules that can initiate brain actions and trigger feelings of fatigue.

   B. Molecules that cross from the blood to the brain to activate the production of nitric oxide.

   C. A type of bacteria that combats abnormalities that arise inside the body.

   D. Independent entities within the body that communicate with nerve fibers.

Part D: Writing

In this part of the test, you will write an essay about the immune system's connection to the nervous and endocrine systems. Write your essay on the following lined pages. Write as neatly as you can. If you want to change a word, cross it out and write a new word above it. When you have finished, you may close the booklet.

Essay

Imagine that your school is developing a new high school textbook to use next year and that it will be written by students. You have been asked to write an essay to be included in the textbook. The editor of the textbook has asked you to include information in your essay that you have learned during the year in biology class as well as information from "The Mind-Body Link" passage in Part A. Specifically, your assignment is to explain:

A. How the immune, endocrine, and nervous systems would work together in a healthy response to an ankle injury.
B. How a breakdown in the communication network (e.g., because of a malfunction in the endocrine system) would affect the healing process in the case of an ankle injury.

In your essay, make sure to:

1. **Explain the most important aspects** of the immune, nervous and endocrine systems.

2. **Organize your ideas logically** so that someone reading your essay can understand the way these three systems interconnect.

3. **Incorporate information** that you learned in class and from the passage in Part A to support these important ideas.

4. **Make connections** between these three systems and the sequence of events related to the body's response to an ankle injury.

5. **Include** an introduction, body, and conclusion in your essay, and write at least five paragraphs.

After you have finished writing, you may want to re-read your essay and make corrections.

APPENDIX E

ILA Rubrics

HOLISTIC RATING

To effectively write an explanation of a biological or scientific process

Score Point	Criteria for Scoring
4	• The response demonstrates in-depth understanding of the relevant and important ideas. • The response very effectively communicates an explanation of the biological and/or scientific process. • The organization enhances the central ideas. • The key concepts are logically organized and paragraphing is correct. • Most or all of the essay's organizational components are strong.
3	• The response includes some of the important ideas related to the topic. • The content's organization is clear and coherent. • The response's order and structure are apparent. • The sequence or cause/effect of processes and facts is logical. • Paragraphing is evident.
2	• The response may include an important idea, part of an idea, or a few facts but does not develop the ideas or deal with the relationships among the ideas. • The response contains misconceptions, inaccurate or irrelevant information. • The content's organization is skeletal. • The response's order and structure are loosely planned. • The sequence or cause/effect of processes or facts is not consistently logical. • Paragraphing is minimally evident.
1	• The response shows no knowledge or understanding of the topic. • The writing is haphazard and disjointed. • The response lacks organization and coherence. • No plan is evident. • The facts may be randomly presented.

BIOLOGY CONTENT

To give an account of how something works or reasons for a phenomenon

Score Point	Criteria for Scoring
4	• The response demonstrates a well-developed understanding and knowledge of the target biology content. • The content is exceptionally clear, focused, and thoroughly explained and elaborated with strong, supportive evidence.
3	• The response demonstrates an adequate understanding and knowledge of the target biology content. • The content is clear, focused, and elaborated with supportive evidence.
2	• The response demonstrates some understanding and knowledge of the target biology content. • The content's main ideas are understandable, but may be overly broad, simplistic, and lack clarity of purpose. • The evidence is insufficient to support the main ideas.
1	• The response demonstrates very little understanding and knowledge of the target biology content. • Minimal content is included. • The response lacks a central purpose. • The response contains little or no detail. • Extraneous materials may be present.

LANGUAGE

To effectively communicate ideas in a written explanation genre

Score Point	Criteria for Scoring
4	• The response demonstrates very good text cohesion through the regular use of sentence structure variety (e.g., information provided in front of the subject that links a sentence to a previous one) • The response demonstrates consistent use of precise and varied words, including technical terms, and expanded word groups to describe the topic. • The tone is impersonal and authoritative with no or minimal speech markers.
3	• The response demonstrates a good level of text cohesion through the use of appropriate language features. • The response demonstrates an adequate use of precise and varied words, including some technical terms, and expanded words groups to describe the topic. • The tone is often impersonal and authoritative though the writing may contain some speech markers and personal references.
2	• The response demonstrates some text cohesion though the ideas are not linked well with appropriate language features. • The sentences are mechanical, repetitive and simplistic. • The response occasionally demonstrates use of precise and varied words but generally the vocabulary is ordinary and there is little expansion of word groups. • The tone may often be informal with regular uses of speech markers and personal references.
1	• The sentences are awkward, choppy, or rambling, and difficult to follow. • The response demonstrates minimal to no text cohesion. • The word usage is simplistic, repetitive, inappropriate, or over used. • There is little to no evidence of expanded word groups. • The tone is usually informal and personal with an overuse of speech markers.

APPENDIX F

Standards Mapping

Meta function	Scoring indicators	General Language Features: word groups	Specific Language Features within general word groups	Genre characteristics of language features	Related California Content Standards in Language Arts: Writing 6th–10th grades
field	Quality of word groups as defined by lexical • Precision • Variety and word group • Expandedness • Genre specific use	1. Noun groups 2. Verb groups 3. Adverbial groups	Noun groups can consist of: main noun, adjectives, embedded clauses, prepositional phrases. Verb groups can consist of: verb, adverbs, prepositional phrase Adverbial groups include: adverbs, subordinate and participial clauses, prepositional phrases	Noun groups: 1a. are often events or happenings instead of personal noun groups 1b. name points to be made (e.g., *There are three reasons that…*) Verb Groups: 2a. frequent action and having/being verbs Adverbial groups: 3a. rank and condense information through use of subordinate clauses	WS 1.2b (6th, 7th, and 8th grades) Develop a topic with supporting details and precise verbs, nouns, and adjectives to paint a visual image in the mind of the reader. WS 1.3 (9th and 10th grade) Use precise language, action verbs, sensory details, appropriate modifiers, and active rather than passive voice. WA 2.3 (9th and 10th grade) Write expository compositions, including analytical essays and research reports: a. Convey information and ideas from primary and secondary sources accurately and coherently d. Anticipate and address readers' potential misunderstandings, biases, and expectations e. Use technical terms and notations correctly.

123

Meta function	Scoring indicators	General language features: structure text	Specific Language Features found within general text structures	Genre Characteristics of Language Features	Related California Content Standards in Language Arts: Writing 6th – 10th grades
mode	Appropriateness of text organization to realize genre, including structures that create: • Balance • Progress • Logic	1. Logical and clear progression of ideas 2. Text connectors 3. Marked themes 4. Thematic progress	1a. Paragraph and sentence structuring of ideas (sentences and paragraphs ordered/grouped by function, e.g. to provide introduction, background, statement, argument, evidence, etc.) 2a. Conjunctions, adverbials, verbs 3a. Adverbial groups in front of subject 4a. Theme of one sentence connected to the theme of a previous sentence, e.g. nominalization	Paragraph and sentence structuring provide: 1a. Presentation of a thesis with supportive evidence 1b. Interpretations and generalizations to account for how and why things are as they are 1c. Logical organization that presents causes and makes judgments 1d. Emphasis on generalization, classification, and categorization of support judgments **Text connectors and marked themes** 2 and 3a. Link one part of the text to the next with cohesive ties, causal conjunctions and markers of contrast, classification, and logical sequence 2 and 3b. Grammatical shifts for moving from general to specific and back again	WA 2.2 (6th grade) Write expository compositions (e.g. description, explanation, comparison and contrast, problem and solution): a. State a thesis or purpose b. Explain the situation c. Follow an organizational pattern appropriate to the type of composition d. Offer persuasive evidence to validate arguments and conclusions as needed. WS 1.0 (6th, 7th, and 8th grade) Students write clear, coherent, and focused essays…Essays contain formal introductions, supporting evidence, and conclusions… WS 1.2c (6th grade) Create multiple paragraph expository compositions that conclude with a detailed summary linked to the purpose of the composition WS 1.3 (6th grade) Use a variety of effective and coherent organizational patterns, including comparison and contrast; organization by categories; and arrangement by spatial order, order of importance, or climactic order. WC 1.1 (6th grade) Use simple, compound, and compound-complex sentences; use effective coordination and subordination of ideas to express complete thoughts. WC 1.2 (6th grade) Identify and properly use indefinite pronouns and present perfect, past perfect, and future perfect verb tenses; ensure that verbs agree with compound subjects. WS 1.1 (7th grade) Create an organizational structure that balances all aspects of the composition and uses effective transitions between sentences to unify important ideas.

(table continues)

124

Meta function	Scoring indicators	General language features: structure text	Specific Language Features found within general text structures	Genre Characteristics of Language Features	Related California Content Standards in Language Arts: Writing 6th – 10th grades
				Thematic progress 4a. Information ranked and condensed through use of clause organization and nominalization	WS 1.2 (7th grade) Support all statements and claims with anecdotes, descriptions, facts and statistics, and specific examples. WS 1.5 (7th and 8th grade) Achieve an effective balance between researched information and original ideas. WC 1.1 (7th and 8th grade) Use correct and varied sentence types and sentence openings to present a lively or effective personal style. WC 1.2 (7th and 8th grade) Identify and use parallelism, including similar grammatical forms, all written discourse to present items in a series and items juxtaposed for emphasis. WC 1.3 (7th and 8th grade) Use subordination, coordination, apposition, and other devices to indicate clearly the relationship between ideas. WS 1.1 (8th grade) Create compositions that establish a controlling impression, have a coherent thesis, and end with a clear and well-supported conclusion. WS 1.2 (8th grade) Establish coherence within and among paragraphs through effective transitions, parallel structures, and similar writing techniques WS 1.3 (8th grade) Support theses or conclusions with analogies, paraphrases, quotations, opinions from authorities, comparisons, and similar devises

125

Meta function	Scoring indicators	Specific language features	Genre Characteristics of Language Features	Related California Content Standards in Language Arts: Writing 6th – 10th grades
tenor	1. Authoritative, formal tone 2. Implicit opinion and evaluation	1a. Passive voice 1b. Third person 1c. No use of speech markers 1d. No address to oneself or audience 2a. Modal auxiliaries 2b. “It” constructions 2c. Precise word choice of nouns, adjectives, verbs, and adverbs to convey evaluation, e.g., *responsible*	2a. and b. Used to present claims as possibilities	WS 1.1 (9th and 10th grade) Establish a controlling impression or coherent thesis that conveys a clear and distinctive perspective on the subject and maintain a consistent tone and focus throughout the piece of writing. WS 1.9 (9th and 10th grade) Revise writing to improve the logic and coherence of the organization and controlling perspective, the precision of word choice, <u>and the tone by taking into consideration the audience, purpose, and formality of the context</u>.

126

APPENDIX G

Assignment Coversheet

Teacher Assignment Collection

Genetics/Physiology

Teacher Name	
School Name	
Type of Biology Class	
Student Grade Span	
Period	

CRESST

National Center for Research on Evaluation, Standards, and Student Testing

ASSIGNMENT COVER SHEET FOR GENETICS/PHYSIOLOGY

1. Reading Material Information

If students read as part of this genetics/physiology assignment, please write the text type (e.g., textbook, journal article), title, author, and reading level of any material students read as part of this assignment. If they received their information from other sources please list them here and specify the type (e.g., film, lecture, laboratory, or investigation).

Text Type	Text Title	Author	Reading Level
a.			
b.			
c.			

2. Reading Comprehension Strategies

If applicable, please write the reading comprehension strategies students used to access the biology content for the genetics/physiology assignment text (e.g., taking notes, reading logs, talking to the text, discussions, or completing graphic organizers). Additionally, if possible, please attach a copy of those reading comprehension tools.

Reading Comprehension Strategy Types
a.
b.
c.

3. Genetics/Physiology Assignment Description

Describe the genetics/physiology assignment in detail. Additionally, if applicable, please attach a copy of the assignment directions you distributed to students.

Learning Goals for Students

What were your *learning* goals for this assignment? Please describe the science skills (and literacy skills, if applicable), biology concepts (and literacy concepts, if applicable), and/or facts you wanted students to learn as a result of completing this genetics/physiology assignment.

4. Description of Support for Student Learning

Describe the support you provided for students in their reading and writing processes (e.g., scaffolding, modeling, explicit instruction, resources, discussion opportunities, etc.). If you differentiated levels of support for different students, please also describe this below. Please attach any support materials you provided to students.

5. Instructional Context

How did this genetics/physiology assignment fit in with your unit, or what you are teaching in your biology/life sciences class this month or this year?

a. How long did students take to complete this assignment? __________

b. Approximately how many assignments like this do you give a year? [ ]

6. Grading Criteria

Please describe your criteria for grading student work for this assignment. If you used a rubric, please attach a copy of the rubric you used to grade student work for this genetics/physiology assignment.

a. If you used a rubric to grade student work for this genetics/physiology assignment, where did this rubric originate? Please check one or more of the following.

[ ] Self
[ ] Students
[ ] Teachers at my school
[ ] District, cluster, or school family
[ ] Published instructional program or teacher's guide
[ ] Other (please describe) ____________________________________

b. Approximately what percentage of the students in your class performed at the following levels for this genetics/physiology assignment?

_____% = Good to Excellent _____% = Adequate _____% = Not Yet Adequate

c. What criteria did you use to decide what was "Medium" student work and what was "High" student work for this genetics/physiology assignment? Please give specific examples from the papers you attach.

APPENDIX H

Observation Protocol

Inside the Biology Classroom
Observation and Analytic Protocol

Observation Date:____________________ Time: Start: ___________ End: ___________

School: ___________________________ District: _________________________________

Teacher:

____

Grade Level(s): ___________ Course Title (if applicable): ________________________

Note:

* Be sure to collect all the instructional materials used in this lesson
* Ask the teacher about the textbook/program/instructional materials, including publisher and title.
* Basic descriptive information (See Part 1, Section A. #1 and #2) should be collected either after or before class by talking to the teachers.

Part 1. The Lesson

Section A. Basic Descriptive Information

I. Teacher

1. Teacher Gender: ___ Male [1]   ___ Female[2]

Teacher Ethnicity: ___ American Indian or Alaskan Native[1]
___ Asian[2]
___ Hispanic or Latino[3]
___ Black or African-American[4]
___ Native Hawaiian or Other Pacific Islander[5]
___ White[6]

2. Number of years teacher has taught prior to this school year:

II. Students

1. Number of students:
   i. Total in class: __________
   ii. Number of males: _____________ number of females: ________________
   iii. For whom English is not their first language: ______
   iv. With learning disabilities: ________
   v. With other special needs: _______

2. Describe the ability level of students in this class compared to the student population in the school. (Check one).
   - ❍ Represent the lower range of ability levels1
   - ❍ Represent the middle range of ability levels2
   - ❍ Represent the higher range of ability levels3
   - ❍ Represent a broad range of ability levels4

III. The Physical Environment

We are defining the physical environment as including:

- Size and "feel" of the room, including what's on the walls;
- State of repair of classroom facilities;
- Appropriateness and flexibility of furniture;
- Availability of running water, electrical outlets, storage space; and
- Availability of equipment and supplies (including calculators and computers).

III. The Physical Environment (continued)

a. Describe the physical environment of this classroom by drawing (include a classroom entrance[s], details of student seating, location of classroom resources such as computers, blackboards, books and so on).

b. Did the physical environment constrain the design and/or implementation of this lesson? (Circle one).

Yes No Don't know *If yes, explain:*

IV. Instructional Materials

a. Which best describes the source of the **instructional materials** upon which this lesson was based? (Check one)

- ❍ Materials designated for this class/course, from a commercially published textbook/program
- ❍ Materials selected or adapted by the teacher, from a non-commercial source
- ❍ Materials developed by the teacher

b. Describe the textbook/program/instructional materials, including publisher and title if applicable (ask the teacher).

Section B. Purpose of the Lesson:

Indicate whether the teacher explicitly conveyed the purpose of this lesson to the class. If yes, how was it conveyed (e.g., through writing on the board, through verbal communication, etc.), and what was it?

Section C. Lesson Ratings

In this part of the form, you are asked to rate a number of key indicators in four different categories, from 1 (not at all) to 5 (to a great extent). You may list any additional indicators you consider important in capturing the essence of this lesson and rate these as well. Use your "Ratings of Key Indicators" to inform your "Synthesis Ratings". It is important to indicate in "Supporting Evidence for Synthesis Ratings" what factors were most influential in determining your synthesis ratings and to give specific examples and/or quotes to illustrate those factors.

Note that any one lesson is not likely to provide evidence for every single indicator; Use 6 (don't know) when there is not enough evidence for you to make a judgment. Use 7 (N/A, not applicable) when you consider the indicator inappropriate, given the purpose and context of the lesson.
This section also includes ratings of the likely impact of instruction and a capsule rating of the quality of the lesson.

I. Instructional Strategies ***A. Ratings of Key Indicators***	Not at all				To a great extent	Don't know	N/A
1. The instructional strategies were consistent with investigative science. (e.g., providing opportunities for students to make predictions, generate hypotheses, use scientific knowledge and understanding to turn ideas into a form that can be investigated, etc.)	1	2	3	4	5	6	7
2. The teacher appeared confident in his/her ability to teach science.	1	2	3	4	5	6	7
3. The teacher's classroom management style/strategies did not distract from the lesson.	1	2	3	4	5	6*	7*
4. The pace of the lesson was appropriate for the developmental levels/needs of the students and the purposes of the lesson.	1	2	3	4	5	6*	7*
5. The teacher adopted formative assessment strategies to gauge the students' level of understanding (e.g., providing feedback during classroom discussion, giving quizzes, analyzing student work, etc.)	1	2	3	4	5	6	7
6. The teacher's questioning strategies were likely to enhance the development of student conceptual understanding/problem solving. (e.g., emphasized higher order questions, appropriately used" wait time," identified prior conceptions and misconceptions).	1	2	3	4	5	6	7
7. The teacher adopted instructional strategies that promote student autonomy (e.g., gradually reducing direct teaching of content knowledge and engaging students in group or individual	1	2	3	4	5	6	7

* We anticipate that these indicators should be rated 1–5 for nearly all lessons. If you rated any of these indicators 6 or 7, please provide an explanation in your supporting evidence below.

B. Synthesis Rating

1	2	3	4	5
Implementation of the lesson not at all reflective of best practice in science education				Implementation of the lesson extremely reflective of best practice in science education

I. Instructional Strategies (continued)

C. Supporting Evidence for Synthesis Rating

Provide a brief description of the nature and quality of this component of the lesson, the rationale for your synthesis rating, and the evidence to support that rating. [If available, be sure to include examples/quotes to illustrate ratings of teacher questioning (A6)].

II. Science Instruction ***A. Ratings of Key Indicators***	Not at all				To a great extent	Don't know	N/A
1. Science was portrayed as a dynamic body of knowledge continually enriched by conjecture, investigation analysis, and/or proof/justification.	1	2	3	4	5	6	7
2. Elements of science abstraction (e.g., symbolic representations, theory building) were included when it was important to do so.	1	2	3	4	5	6	7
3. Appropriate connections were made to other areas of science, to other disciplines, and/or to real-world contexts.	1	2	3	4	5	6	7
4. The degree of "sense-making" of science content within this lesson was appropriate for the purposes of the lesson (e.g., purposeful use of instructional formats and activities).	1	2	3	4	5	6	7
5. Various forms of objects (e.g., graphs, models, visual aids, experiment equipment, etc.) were incorporated in this lesson to enhance students' understanding of science concepts.	1	2	3	4	5	6	7

B. Synthesis Rating

1	2	3	4	5
Instruction in this science lesson is not at all reflective of best practices in science education				Instruction in this science lesson is extremely reflective of best practices in science education

II. Science Content (continued)

C. Supporting Evidence for Synthesis Rating

Provide a brief description of the nature and quality of this component of the lesson, the rationale for your synthesis rating, and the evidence to support that rating. (If available, be sure to include examples/quotes to illustrate ratings of quality of content (A1, A2, A3), intellectual engagement (A4), and nature of "sense-making" (A9).

III. Literacy Content *A. Ratings of Key Indicators*	Not at all				To a great extent	Don't know	N/A
1. Literacy instruction (reading and/or writing) was incorporated effectively in this lesson to support science learning goals.	1	2	3	4	5	6	7
2. Adequate time was provided for teacher- and peer-supported in-class reading and/or discussion of science materials.	1	2	3	4	5	6	7
3. Supplementary reading material beside a textbook (e.g., newspaper articles, print-outs from a website, etc.) was used in this lesson to accomplish the purpose and learning goals of this science lesson.	1	2	3	4	5	6	7
4. The teacher engaged students in a writing activity to provide them with an opportunity to extend newly acquired scientific knowledge.	1	2	3	4	5	6	7
5. The values and importance of scientific literacy (i.e., language as a medium for communicating scientific content and ideas) were conveyed to students either explicitly or implicitly.	1	2	3	4	5	6	7
6. The teacher employed various scaffolding strategies in order to elicit extended oral discourse from students (e.g., paraphrasing, asking follow-up questions, etc.).	1	2	3	4	5	6	7
7. Additional linguistic support (e.g., help with translation, modified worksheets, etc.) was provided in consideration of English Language Learners or students with special needs.	1	2	3	4	5	6	7
8. Students' linguistic and cultural practices, knowledge and experiences were invited into the classroom and leveraged for students' literacy learning	1	2	3	4	5	6	7

B. Synthesis Rating

1	2	3	4	5
Literacy content of lesson not at all reflective of best practice in linguistic support in science content area.				Literacy content of lesson extremely reflective of best practice in linguistic support in science content area.

III. Literacy Content (continued)

C. Supporting Evidence for Synthesis Rating

Provide a brief description of the nature and quality of this component of the lesson, the rationale for your synthesis rating, and the evidence to support that rating. [If available, be sure to include examples/quotes to illustrate literacy events (any reading or writing activity) as well as instructional strategies used in the instruction of scientific literacy].

IV. Classroom Culture ***A. Ratings of Key Indicators***	Not at all				To a great extent	Don't know	N/A
1. Active participation of all was encouraged and valued.	1	2	3	4	5	6*	7*
2. There was a climate of respect for students' ideas, questions, and contributions.	1	2	3	4	5	6*	7*
3. Interactions reflected collegial working relationships among students (e.g., students worked together, talked with each other about the lesson).	1	2	3	4	5	6	7
4. Interactions reflected collaborative working relationships between teacher and students	1	2	3	4	5	6*	7*
5. The climate of the lesson encouraged students to generate ideas, questions, conjectures, and/ or propositions	1	2	3	4	5	6	7

* We anticipate that these indicators should be rated 1–5 for nearly all lessons. If you rated any of these indicators 6 or 7, please provide an explanation in your supporting evidence below.

B. Synthesis Rating

1	2	3	4	5
Classroom culture interfered with student learning				Classroom culture facilitated the learning of all students

IV. Classroom Culture (continued)

C. Supporting Evidence for Synthesis Rating

Provide a brief description of the nature and quality of this component of the lesson, the rationale for your synthesis rating, and the evidence to support that rating. [If available, be sure to include examples/quotes to illustrate ratings of active participation (Al), climate of respect (A2), and intellectual rigor (A6). While direct evidence that reflects particular sensitivity or insensitivity toward student diversity is not often observed, we would like you to document any examples you do see].

Section D. Lesson Arrangements and Activities

In question 1 of this section, please divide the total duration of the lesson into instructional and non-instructional time. In question 2, make your estimates based only on the *instructional time* of the lesson.

1. Approximately how many minutes during the lesson were spent:

   a. On instructional activities? _____minutes
   b. On housekeeping unrelated to the lesson/interruptions/other non-instructional activities?_____minutes

Describe non-instructional activities:

   c. Check here if the lesson included a major interruption (e.g., fire drill, assembly, shortened class period): ❑

2. Considering only the *instructional time* of the lesson (listed in 1a above), approximately what percent of this time was spent in each of the following arrangements?

   a. Whole class

   b. Pairs/small groups

   c. Individuals

   ______________ 100

Part 2. Nonverbal Immediacy Score-Observer

In this part you are being asked to rate the frequency of the teacher's nonverbal behavior while talking to the students. Nonverbal Immediacy is the nonverbal communication behavior (e.g. hand gestures, eye contact, smiling, body language, and tone of voice) of teachers and has been linked to effective teacher communication, and in particular, its impact on the affective learning of students. Affective learning addresses a learner's emotions towards learning experiences. A learner's attitudes, interest, attention, awareness, and values are demonstrated by affective behaviors.

The following statements describe the ways teachers behave while talking with or to students. Please indicate in the space at the left of each item the degree to which you believe the statement applies to the teacher. Please use the following 5-point scale:

1 = never; 2 = rarely; 3 = occasionally; 4 = often; 5 = very often

_____ 1. The teacher uses his or her hands and arms to gesture while talking to students.
_____ 2. The teacher touches others on the shoulder or arm while talking to them.
_____ 3. The teacher uses a monotone or dull voice while talking to students.
_____ 4. The teacher looks over or away from others while talking to them.
_____ 5. The teacher moves away from others when touched during talk.
_____ 6. The teacher has a relaxed body position when talking to students.
_____ 7. The teacher frowns while talking to students.
_____ 8. The teacher avoids eye contact while talking to students.
_____ 9. The teacher has a tense body position while talking to students.
_____ 10. The teacher sits or stands close to students while talking to them.
_____ 11. The teacher's voice is monotonous or dull when he or she talks to students.
_____ 12. The teacher uses a variety of vocal expressions when talking to students.
_____ 13. The teacher gestures when talking to students.
_____ 14. The teacher is animated when talking to students.
_____ 15. The teacher has a bland facial expression when talking to students.
_____ 16. The teacher moves closer to students when talking to them.
_____ 17. The teacher looks directly at students while talking to them.
_____ 18. The teacher is stiff when talking to students.
_____ 19. The teacher has a lot of vocal variety when he or she talks to students.
_____ 20. The teacher avoids gesturing while he or she is talking to students.
_____ 21. The teacher leans toward students when talking to them.
_____ 22. The teacher maintains eye contact with students when talking to them.
_____ 23. The teacher tries not to sit or stand close to students when talking to them.
_____ 24. The teacher leans away from students when talking to them.
_____ 25. The teacher smiles when talking to students.
_____ 26. The teacher avoids touching students when talking to them.

Scoring for NIS-O:
Step 1: Start with a score of 78. Add the scores from the following items: 1, 2, 6, 10, 12, 13, 14, 16, 17, 19, 21, 22, and 25
Step 2: Add the scores from the following items: 3, 4, 5, 7, 8, 9, 11, 15, 18, 20, 23, 24, and 26
Total score = Step 1 minus Step 2