2009-2010

REVIEW OF THE BIOLOGY 1/APPLIED BIOLOGY 2 END-OF-COURSE FIELD TEST

SC EDUCATION
OVERSIGHT COMMITTEE

PO Box 11867 | 227 Blatt Building | Columbia SC 29211 | WWW.SCEOC.ORG

Review of the Biology 1/Applied Biology 2 End of Course Test
Executive Summary

This report summarizes the results from the review of the 2009 Biology 1/ Applied Biology 2 end of course field test and item bank by the Education Oversight Committee (EOC) pursuant to Section 59-18-320 (A) of the Education Accountability Act. The reviews of the items and technical data were conducted in April 2010 with the assistance of 12 South Carolina high school biology educators who served as expert judges evaluating the content tested and the levels of thinking demanded by the Biology 1/Applied Biology 2 items provided to the EOC by the SC Department of Education (SCDE) for the review, and the assistance of four South Carolina technical experts who evaluated the statistical characteristics of the 2009 Biology 1/Applied Biology 2 field test following the test administration in spring 2009. The criteria for the alignment review were based on a methodology used by most states which was developed by Norman Webb and colleagues at the University of Wisconsin with the assistance of the Council of Chief State School Officers, and the criteria for technical review were based on generally accepted statistical values for use in large-scale testing programs.

The results from the review of the alignment of the Biology 1/Applied Biology 2 end of course test items reveal that they are well aligned to the state academic course standards. The field test items analyzed met all four of the Webb methodology criteria for test alignment:

1. Categorical Concurrence: there were more than 6 items assessing each standard and the proportions of items assessing the standards were similar to the proportions of items for each standard listed in the test Blueprint;
2. Depth-of-Knowledge: 52% of the items assessed at or above the cognitive process levels specified for the academic standard indicators;
3. Range-of-Knowledge: there were from 5 to 25 items identified for assessing each indicator, suggesting that sufficient items were available for each indicator to construct more than one test form which would meet the specifications of the test Blueprint;
4. Balance-of-Representation: the indexes of balanced representation for each standard met the criteria for tests which assess the content in a comprehensive and balanced manner, without under- or over-representing some objectives compared to others.

The technical review concluded that, while the field test was difficult, the results were found to be within an acceptable range for a field test. The test reliabilities were also within an acceptable range for field tests and the item statistics for most of the items were also within acceptable ranges. However, some items were flagged for being exceptionally difficult and some items were flagged for poor discrimination between students having overall high scores on the rest of the items on the field test and students having overall low scores. There were indications from the student survey responses that the test may have taken two hours or more for administration, which may cause conflicts in high school class schedules. Finally, the population of students taking the Biology 1/Applied Biology 2 end of course test is expected to change when the course becomes required for all students such that more students enrolled in Applied Biology 2 will be taking the test than were evident in the field test sample.

Recommendations

1. The items flagged for out-of-range item statistics should be reviewed to identify the reasons for the observed item statistics and, if necessary, revised before use in future test forms.

2. The Biology 1/Applied Biology 2 end of course test is not a “timed” test: students can work on the test as long as needed to complete it. The length of time needed to administer the test which was reported from the field test survey (90% of the students reporting 2 hours or more on six of the nine forms) should be reviewed and if verified, the SCDE should consult with school district personnel regarding ways to address the scheduling needed to provide appropriate time for administering the test.

3. Since the student sample participating in the Biology 1/Applied Biology 2 field test may not be fully representative of the student cohorts required to take the assessment in the future because of federal requirements, the scaling and standard setting of the Biology 1/Applied Biology 2 test should be revisited when the test becomes required and more representative information becomes available from the broader population tested. The anticipated increase in the number of students enrolled in Applied Biology 2 when the test becomes required may result in lower overall test performance and may change the distribution of scores enough to warrant review and possible adjustment of the scaling and performance standards.

Review of the Biology 1/Applied Biology End of Course Test

Introduction

This report summarizes the results from studies of the Biology 1/Applied Biology 2 end of course field test administered in Spring 2009. The report includes the review of the alignment of the Biology 1/Appied Biology 2 item bank to the South Carolina Biology academic standards and the review of the technical data resulting from the Spring 2009 field test. The studies were conducted under the auspices of the Education Oversight Committee (EOC) as part of its responsibilities listed in the Education Accountability Act of 1998 (EAA):

> Section 59-18-320. (A) After the first statewide field test of the assessment program in each of the four academic areas, and after the field tests of the end of course assessments of high school credit courses, the Education Oversight Committee, established in Section 59-6-10, will review the state assessment program and the course assessments for alignment with the state standards, level of difficulty and validity, and for the ability to differentiate levels of achievement, and will make recommendations for needed changes, if any. The review will be provided to the State Board of Education, the State Department of Education, the Governor, the Senate Education Committee, and the House Education and Public Works Committee as soon as feasible after the field tests. The Department of Education will then report to the Education Oversight Committee no later than one month after receiving the reports on the changes made to the assessments to comply with the recommendations.
>
> (C) After review and approval by the Education Oversight Committee, the end of course assessments of high school credit courses will be administered to all public school students as they complete each course.

The report describes the Biology 1/Applied Biology 2 end of course test, describes the studies conducted for this review, presents the findings from the studies, and makes recommendations regarding the assessments. Two studies were conducted for this review. One study was designed to determine the alignment of the items in the Biology 1/Applied Biology 2 item bank to the state academic course standards. A second study was conducted focusing on the technical characteristics of the items and tests, and was conducted in April 2010 subsequent to the administration of the field test.

<u>Biology 1/Applied Biology 2 End of Course Test and State Biology Course Standards</u>

The Biology 1/Applied Biology 2 end of course test is intended to assess student proficiency of the Biology course standards at the end of the student's course work in either Biology 1 or Applied Biology 2. Upon approval of the test by the EOC the results will be used for several purposes: to inform and improve science instruction and learning; to address student accountability for their learning by counting as 20% in the calculation of the student course grade; to address school and district accountability by inclusion in the calculation of school and district report card ratings and by inclusion in the federal ESEA accountability system.

The 2009 Biology 1/Applied Biology 2 field test marks the return of end of course testing in the high school Biology course. A Biology 1/Applied Biology 2 end of course test, based on science standards adopted in 2000, was administered from 2004 through 2006. The administration of this test was ended in 2007 when the Physical Science end of course test was selected to meet federal No Child Left Behind accountability requirements. The Biology 1/Applied Biology 2 field test reviewed in this report is based on the 2005 state science academic standards. It is expected that, upon adoption, the Biology 1/Applied Biology 2 end of course test will replace the Physical Science end of course test to meet federal accountability requirements.

The general design for the Biology 1/Applied Biology 2 end of course test is outlined in the test blueprint, obtained from the South Carolina Department of Education (SCDE) web site and attached in Appendix A. The blueprint helps to clarify to educators how many items on the test assess each of the Biology academic standards. The test blueprint lists the total number of items on the test and provides a guideline as to the numbers of items assessing each of the academic standards. The Biology 1/Applied Biology 2 blueprint indicates that the six standards will have a range of 9 to 12 items each, ranging from standard B-2 with 9-10 items and standard B-4 with 10-12 items allocated. All of the items on the test are multiple choice in format and count one point each in the total test score if correct and zero points if incorrect. Each test will consist of 60 items, yielding a range of possible raw scores on the test from 0 to 60.

The Biology 1/Applied Biology 2 end of course test was constructed based on the expectations for student learning described in the SC Academic Standards for Biology (2005). The academic standards document describes what students are expected to know and be able to do in the subject area upon completion of the course. The test items are written based on the end-of-grade expectations for achievement listed in the academic standards document.

The standards in all the subject areas are written following the same format: the student expectations are listed for each grade level organized into "standards," which are further defined by "indicators" subsumed under each standard. Standards are defined as, "Statements of the most important and consensually determined expectation for student learning in a particular discipline," (SC Academic Standards for English Language Arts, 2008, p. 2) and indicators are defined as, "statements of the specific cognitive processes (expressed in the main verbs) and the content knowledge and skills that students must demonstrate in order to meet the grade-level or high school core area standard" (SC Academic Standards for Science, 2005, p. 7). Indicators in the SC standards documents can be thought of as analogous to "objectives," a term often used in descriptions of educational curriculum goals. The same numbering system is used for all subjects and grades: "B-1.1" indicates Biology Standard 1, Indicator 1. Each test item is written at a level to address a specific, primary indicator, although some items may address secondary indicators as well.

Study of Alignment of Biology 1/Applied Biology 2 End of Course Item Bank to State Academic Standards

The study of the alignment between test items and the academic standards is intended to address the overall question, "Are the tests used to evaluate student proficiency and progress in meeting the expectations for learning specified in the state academic standards actually testing those academic standards?" To address this overall question, the American Educational Research Association (AERA) proposed several related questions to be addressed to determine whether a test is aligned with the academic standards:

- "Does the test's content match the content (topics and skills) in the standards? In other words, each test item should correspond to an objective in the standards. Similarly, key ideas in the standards should appear on the tests.
- Do the tests and standards cover a comparable "range" or breadth of knowledge, and is there an appropriate "balance" of knowledge across the standards? Alignment studies look at whether a test fairly and effectively samples across the range of objectives described in a state's standards instead of focusing on only a few objectives or disproportionately sampling students' knowledge of some objectives but not others.
- Does the level of cognitive demand or challenge called for in the standards match that required for students to do well on the assessment? For example, if the standards

require students to synthesize information and explain their thinking, but the test items only ask students to recall facts, the standards and the tests would not be well aligned.

- Does the test avoid adding material that is irrelevant to the standard supposedly being assessed? For example, a test item may have an inappropriate "source of challenge," requiring a student to read and understand a long passage about space travel, when it is seeking to measure a student's knowledge of how to estimate distances and travel times." (AERA, 2003, p. 2)

Several models have been used by states to analyze their tests' alignments to their academic standards to address these questions. The Council of Chief State School Officers (CCSSO) has collaborated with university researchers and state departments of education for several years to develop methodologies for evaluating test alignment. The CCSSO and states have participated extensively in the development of two alignment evaluation models, the Webb model and the Surveys of Enacted Curriculum (SEC), both developed at the University of Wisconsin Center for Educational Research. Both the Webb and SEC models use information from the academic standards and the tests, but the SEC model also requires the collection of data on the implementation of curriculum in schools and classrooms based on surveys from teachers describing their implementation of curriculum content over the course of the school year (CCSSO, 2002). Since resources were not available to conduct the alignment analysis of the Biology 1/Applied Biology 2 end of course test following the SEC methodology, and since the Webb methodology has been used in more than ten states and has been used for evaluating the alignments in language arts, mathematics, science, and social studies, the Biology 1/Applied Biology 2 alignment has been conducted using the Webb methodology as described in the literature.

The CCSSO describes the Webb alignment methodology as follows:

> "The model developed by Norman Webb provides a reliable set of procedures and criteria for conducting alignment analysis studies, which combine qualitative expert judgments and quantified coding and analysis of standards and assessments. The product of the analysis is a set of statistics for each standard and grade on the degree of intersection, or alignment, between the content embedded in state content standards and the content in state assessments." (CCSSO, 2002, p. 2)

There are four measures calculated in the Webb methodology, with criteria for acceptable levels for each measure. The four measures and the questions they address are:

1. Categorical Concurrence: Are the test items and the academic standards covering the same content?
2. Depth-of-Knowledge Consistency: Are the test items and the academic standards at the same level of cognitive complexity, or is there a mismatch between the complexity of thinking demanded in the standards and the complexity of thinking elicited by the items?
3. Range-of-Knowledge Correspondence: Does the breadth of knowledge and skills assessed by the items match the breadth of knowledge and skills defined in the academic standards; e.g., are most of the indicators in a standard assessed or only a few?
4. Balance of Representation: Are some objectives or indicators in a standard tested while others are not, or are the numbers of items assessing each indicator proportional to the number of indicators in the standard?

Each of the Webb measures and criteria for acceptable values will be described in more detail in the following sections of the report when the results of the analyses are reported.

The Webb measures (as do all the other measures of alignment) require data from expert judges on the content tested by each test item and on the complexity of cognitive processing required to successfully answer each item. These data were collected for the alignment analysis April 22, 2010, when 12 educators (high school Biology teachers and district science curriculum specialists from public school districts) met in Columbia, SC to review the items developed for the Biology 1/Applied Biology 2 end of course test. The reviewers were nominated, based on their expertise and experience teaching their subject area, by school district Superintendents during March 2010. All South Carolina school district superintendents were solicited by EOC staff for nominations to the item review committees, and nominations were received from forty-one public school districts and the SC Public Charter School District. The members of the review committee were chosen based on size and geographic location of district to ensure representation from all areas of the state. The members of the committee reviewing the Biology 1/Applied Biology 2 test items for the alignment study are listed in Appendix B.

Paper copies of the Biology 1/Applied Biology 2 items were provided to EOC staff for the study by the SCDE Office of Assessment in March 2010, along with Excel files containing item ID numbers and other information describing the items. The items reviewed included 285 Biology 1/Applied Biology 2 items field tested in Spring 2009 and an additional 208 items scheduled to be field tested in Spring 2010, for a total of 493 items. The items were sorted into four groups for review so that each of the four groups of items had similar numbers of previously field tested items from 2009 and of items scheduled for field testing in 2010. The review panel was also divided into four groups of three panelists each, with one of the four sets of items to be reviewed assigned to each group of reviewers. It was expected that each reviewer would be able to review a maximum of one hundred twenty-five items during the time allocated for the review meeting, so the items were grouped into review booklets of copies of the items limited to one hundred twenty-five items. The actual maximum number of items in a group was one hundred twenty-five and the minimum number of items in a group was one hundred-twenty. Each item was reviewed by three judges.

At the meeting to review the items the 12 committee members were first assembled as a whole group for introduction and discussion of the purpose for their review, the process they were to follow for the review, state test security statutes and regulations, and to answer questions. Each committee member signed an oath of security regarding the content of the test items they were to review. At the whole-group meeting the process for the judges to review the items was presented as a series of tasks to be accomplished. The data collection system was also described during this session. The tasks the committee members were asked to complete for all of the items they were assigned to review were:

TASK 1: To determine the degree of alignment between the Biology test items and the South Carolina Academic Standards.

The purpose of this activity is to determine the degree to which Biology 1/Applied Biology 2 assessment items match the curriculum standards. Each analyst should work on this task individually. The task will be accomplished in several steps:

<u>Step 1</u> - Read and answer the test question. As you are doing so, reflect on the kinds of knowledge and skills needed to correctly answer the question and on the level of cognitive challenge the question presents to students.

Step 2 - Review the standards document to identify the standard(s) and indicator(s) you believe the item best addresses.

Step 3 - Record the standard(s)/indicator(s) you believe the item is addressing in the space provided. Use the numbering system in the standards document (e. g., B-3.1, B-6.2, etc.) to identify standard(s) and indicator(s). If you identify more than one standard, CIRCLE the standard you believe is the primary one addressed.

TASK 2: To identify the level of cognitive demands made by the item which must be met to correctly answer it.

The purpose of this task is to make a judgment regarding the knowledge dimension and cognitive process for each test question. Refer to the document, "A Taxonomy for Teaching, Learning, and Assessing" (Revised Bloom's Taxonomy).

The cognitive processes are:

1. Remember
2. Understand
3. Apply
4. Analyze
5. Evaluate
6. Create

The knowledge dimensions are:

A. Factual knowledge
B. Conceptual knowledge
C. Procedural knowledge
D. Metacognitive knowledge

Based on your reading of the question, identify the cognitive process and the knowledge dimension needed to correctly answer the item and record the process and dimension in the space provided. Record the cognitive process and the knowledge dimension in the spaces provided (e.g., 3 B, 1 C, etc.).

Following group practice exercises identifying standards and cognitive processes and recording the data using two sample items for review and discussion, the judges were assigned to the four groups. The committee members were provided notebooks containing the items they were to review, data collection forms, and copies of the state Biology academic standards document. The data collection sheets were collected by EOC personnel for analysis and the data were transcribed to Excel files and have provided the data for the alignment analyses in this report.

The analyses reported in this report for the Categorical Concurrence, Range-of Knowledge Correspondence, and Balance-of-Representation alignment measures are based on the reviewers' coding of the grade level, standard(s), and indicator(s) assessed by each item. Reviewers were asked to indicate the primary and secondary standard(s) and indicators assessed if they believed more than one standard or indicator was being assessed by the item. Some items may have been written to assess more than one indicator and reviewers, acting independently, may have identified different standards or indicators for an item. Thus some items are counted more than once, once for each standard/indicator identified, in the tables reported in the following analyses.

The percentage of exact agreement among the reviewers on the primary *standards* assessed by the items was 96.6%, while the percentage of exact agreement among the reviewers on the primary *indicators* assessed by the items was 75.3%. One hundred percent exact agreement on standards and indicators was not expected because some items may have been written to assess more than one standard or indicator. Further, the exact agreement statistic was based on the extent to which all three panelists agreed on the primary standard or indicator assessed by the items; the three panelists reviewing each item may each have assigned the same two standards or indicators to it, but differed in which of the two standards or indicators was primary. When the data were re-examined to assess the extent to which two of the three panelists agreed on the primary *indicator* assessed by the items, it was determined that two of the three panelists chose the same primary indicator for 97.8% of the items.

The reviewer agreement data suggest that most of the Biology 1/Applied Biology 2 items were written to primarily assess only one standard and indicator rather than multiple standards and indicators. The interpretation of the test results is more clear and accurate if the items on the tests are each focused on a single standard and indicator rather than on multiple standards and indicators. One cannot determine with assurance which of the standards and indicators a student does not know if the student gets an item assessing multiple standards and indicators wrong.

Categorical Concurrence

Background and Methodology – Categorical Concurrence

The categorical concurrence measure is intended to address the question, "Is all the content in the academic standards assessed by the test items?" The Webb definition of categorical concurrence is:

> "The criterion of categorical concurrence between standards and assessments is met if the same or consistent categories appear in both documents." Webb, Horton, and O'Neal, 2002, p. 4)

The data for this measure came from the identification by the alignment review committees of the primary standard and indicator measured by each item. Since each item was reviewed by more than one reviewer who independently identified the standard and indicator he or she thought was being measured by the item, and since at least some of the items may have been written to assess both a primary and a secondary standard or indicator, an item may have been assigned to more than one standard and would be duplicate-counted in the compilation of the data for the tables below displaying the analyses.

The criterion for meeting the Webb measure of categorical concurrence is that there should be a minimum of six items assessing a standard on a test for alignment to be deemed acceptable. There is no published criterion for how many items in an item bank should be measuring each standard, so the Webb criterion was used to evaluate categorical concurrence for this analysis. However, one should keep in mind that the Webb criterion for a test is probably too low for an item bank because new items are selected from an item bank each year to generate new tests. If the item bank has only six items for a standard and all six items must be used in the test form to provide a reliable score for the standard, all the items would need to be re-used each year until more items are developed. The criterion for an item bank is also problematic because the standard may have more than six indicators, and to adequately measure students' knowledge and skill in the standard they should be tested on every indicator if possible, or a representative

sample of indicators if there are too many to assess each year. If there are six items in the bank for a standard and eight indicators, for example, at least two indicators would never be tested.

Findings – Categorical Concurrence

The results of the evaluation of categorical concurrence are displayed in Table 1 A more detailed table listing the numbers of items identified measuring each indicator is available in Appendix C.

The first finding to note is that there are more than six items assessing each standard, thus meeting the Webb criterion.

A second observation from the tables is that the percentages of items assessing individual standards in the item bank generally reflect the relative weight for the standard in the test blueprint. For example, the Biology 1/Applied Biology 2 Blueprint specifies that the overall percentage of items assessing Standard 1 (Scientific Inquiry) on the test should range from 16.7% to 18.3%, and the reviewers identified 19.0% of the items as assessing Standard 1. Thus it appears that there are sufficient items available for each of the standards tested to develop more than one form for future use.

Table 1
Items Aligned to Academic Standards
2009 Biology 1/Applied Biology 2 End of Course Field Test Data

Standard	# Indicators in Standard	# Items Identified Assessing Standard*	% of All Items Assessing Standard	Range of % of Items Assessing Standard from Blueprint
1: Scientific Inquiry	9	119	19.0	16.7-18.3
2: Structure and Function of Cells and Their Organelles	8	94	15.0	15.0-16.7
3: Energy Flow Within and Between Living Systems	6	91	14.6	15.0-18.3
4: Molecular Basis of Heredity	9	125	20.0	16.7-20.0
5: Biological Evolution and the Diversity of Life	7	99	15.8	15.0-18.3
6: Interrelationships Among Organisms and Biotic and Abiotic Components of Their Environments	6	97	15.5	15.0-18.3

* An item may have been counted more than once if different judges assigned different standards or indicators to the item.

Depth-of-Knowledge Consistency

Background and Methodology – Depth-of-Knowledge

Alignment of assessment items and the content of the academic standards is essential but not sufficient to ensure that what is tested reflects expectations for the levels of student understanding of that content inherent in the standards. Does an academic standard expect that students remember or recognize key terms or ideas in a content area, or does it rather expect students to be able to compare or differentiate the attributes of those key terms or ideas? The test items assessing the standard should be written so that students need to apply a cognitive process at the level required by the standard if they are to answer the item successfully. When the expectations for cognitive processing expressed in a standard are reflected in test items assessing that standard then the depth of knowledge needed to correctly answer the items is said to be consistent with the standard. This criterion for evaluating the alignment between tests and academic standards is termed "Depth-of-Knowledge Consistency" and is defined as:

> "Depth-of-knowledge consistency between standards and an assessment indicates alignment if what is elicited from students on the assessment is as demanding cognitively as what students are expected to know and do as stated in the standards." Webb, Horton, and O'Neal, 2002, p. 5)

The depth-of-knowledge criterion asks the question, "Do the levels of thinking elicited by the test items match the levels of thinking specified in the academic standards?" To answer this question required two kinds of information: the cognitive processes specified in the academic standards and the cognitive processes elicited by the test items. One of the attributes of South Carolina's academic standards is that the cognitive levels of the indicators within the standards for all subjects are specified based on the cognitive processes described and defined in the revised Bloom's taxonomy of educational objectives (Anderson & Krathwohl, 2001). This widely-used resource describing the cognitive processes underlying educational objectives lists six main categories of cognitive processing:

1. Remember
2. Understand
3. Apply
4. Analyze
5. Evaluate
6. Create.

Within each category is a list of specific cognitive processes defining and giving examples to explain each broad category (Table 2). For example, the Remember category has two components, Recognizing and Recalling. In the process of writing the academic standards, the specific cognitive processing terms in the revised Bloom's taxonomy were converted to transitive verbs generally placed at the beginning of the sentence defining the indicator. For example, "recalling," which is a cognitive process within the category "Remember," was changed to "recall" in this example Biology indicator:

> Recall the three major tenets of cell theory (all living things are composed of one or more cells; cells are the basic units of structure and function in living things; and all presently existing cells arose from previously existing cells). (SC Biology Standards, Standard 2, Indicator 1).

Table 2: The Cognitive Process Dimension		
CATEGORIES & COGNITIVE PROCESSES	**ALTERNATIVE NAMES**	**DEFINITIONS AND EXAMPLES**
1. REMEMBER—Retrieve relevant knowledge from long-term memory		
RECOGNIZING	Identifying	Locating knowledge in long-term memory that is consistent with presented material (e.g., Recognize the dates of important events in United States history)
RECALLING	Retrieving	Retrieving relevant knowledge from long-term memory (e.g., Recall the dates of important events in United States history)
2. UNDERSTAND—Construct meaning from instructional messages, including oral, written, and graphic communication		
INTERPRETING	Clarifying, paraphrasing, representing, translating	Changing from one form of representation (e.g., numerical) to another (e.g., verbal) (e.g., Paraphrase important speeches and documents)
EXEMPLIFYING	Illustrating, instantiating	Finding a specific example or illustration of a concept or principle (e.g., Give examples of various artistic painting styles)
CLASSIFYING	Categorizing, subsuming	Determining that something belongs to a category (e.g., Classify observed or described cases of mental disorders)
SUMMARIZING	Abstracting, generalizing	Abstracting a general theme or major point(s) (e.g., Write a short summary of events portrayed on a videotape)
INFERRING	Concluding, extrapolating, interpolating, predicting	Drawing a logical conclusion from presented information (e.g., In learning a foreign language, infer grammatical principles from examples)
COMPARING	Contrasting, mapping, matching	Detecting correspondences between two ideas, objects, and the like (e.g., Compare historical events to contemporary situations)
EXPLAINING	Constructing models	Constructing a cause-and-effect model of a system (e.g., Explain the causes of important 18th Century events in France)

Table 2: The Cognitive Process Dimension		
CATEGORIES & COGNITIVE PROCESSES	**ALTERNATIVE NAMES**	**DEFINITIONS AND EXAMPLES**
3. APPLY—Carry out or use a procedure in a given situation		
EXECUTING	Carrying out	Applying a procedure to a familiar task (e.g., Divide one whole number by another whole number, both with multiple digits)
IMPLEMENTING	Using	Applying a procedure to an unfamiliar task (e.g., Use Newton's Second Law in situations in which it is appropriate)
4. ANALYZE—Break material into its constituent parts and determine how the parts relate to one another and to an overall structure or purpose		
DIFFERENTIATING	Discriminating, distinguishing, focusing, selecting	Distinguishing relevant from irrelevant parts or important from unimportant parts of presented material (e.g., Distinguish between relevant and irrelevant numbers in a mathematical word problem)
ORGANIZING	Finding coherence, integrating, outlining, parsing, structuring	Determining how elements fit or function within a structure (e.g., Structure evidence in a historical description into evidence for and against a particular historical explanation)
ATTRIBUTING	Deconstructing	Determine a point of view, bias, values, or intent underlying presented material (e.g., Determine the point of view of the author of an essay in terms of his or her political perspective)
5. EVALUATE—Make judgments based on criteria and standards		
CHECKING	Coordinating, detecting, monitoring, testing	Detecting inconsistencies or fallacies within a process or product; determining whether a process or product has internal consistency; detecting the effectiveness of a procedure as it is being implemented (e.g., Determine if a scientist's conclusions follow from observed data)
CRITIQUING	Judging	Detecting inconsistencies between a product and external criteria, determining whether a product has external consistency; detecting the appropriateness of a procedure for a given problem (e.g., Judge which of two methods is the best way to solve a given problem)

6. CREATE—Put elements together to form a coherent or functional whole; reorganize elements into a new pattern or structure		
GENERATING	Hypothesizing	Coming up with alternative hypotheses based on criteria (e.g., Generate hypotheses to account for an observed phenomenon)
PLANNING	Designing	Devising a procedure for accomplishing some task (e.g., Plan a research paper on a given historical topic)
PRODUCING	Constructing	Inventing a product (e.g., Build habitats for a specific purpose)

From Lorin W. Anderson and David R. Krathwohl, *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Educational Objectives,* © 2001. Published by Allyn and Bacon, Boston, MA. © 2001 by Pearson Education. Reprinted by permission of the publisher.

The cognitive process descriptions in the revised Bloom's taxonomy have been widely distributed to South Carolina educators to raise awareness among educators of the importance of considering the levels of thinking required as well as the content in a subject when planning lessons for their students. For example, statements such as the following from the SC Science standards document are also found in the other content areas:

> "The main verbs in the indicators are taxonomic—that is, they identify specific aspects of the cognitive process as described in the revised Bloom's taxonomy, which is included in this standards document in appendix B. Use of this new taxonomy will allow teachers to identify the kind of content (knowledge) addressed in the indicators (as factual, conceptual, procedural, or metacognitive) and will help teachers to align their lessons with both the content and the cognitive process identified in the indicators.
>
> Many of the indicators in science address conceptual knowledge and fall under the second category of cognitive processing, *understanding*, which fosters transfer and meaningful learning rather than rote learning and memorization. These revised science standards also contain some indicators that require students to *analyze* or *evaluate* data and/or the results of investigations so that they must use understanding as they demonstrate even more cognitively complex learning." (SC Science Standards, 2005, p. 7)

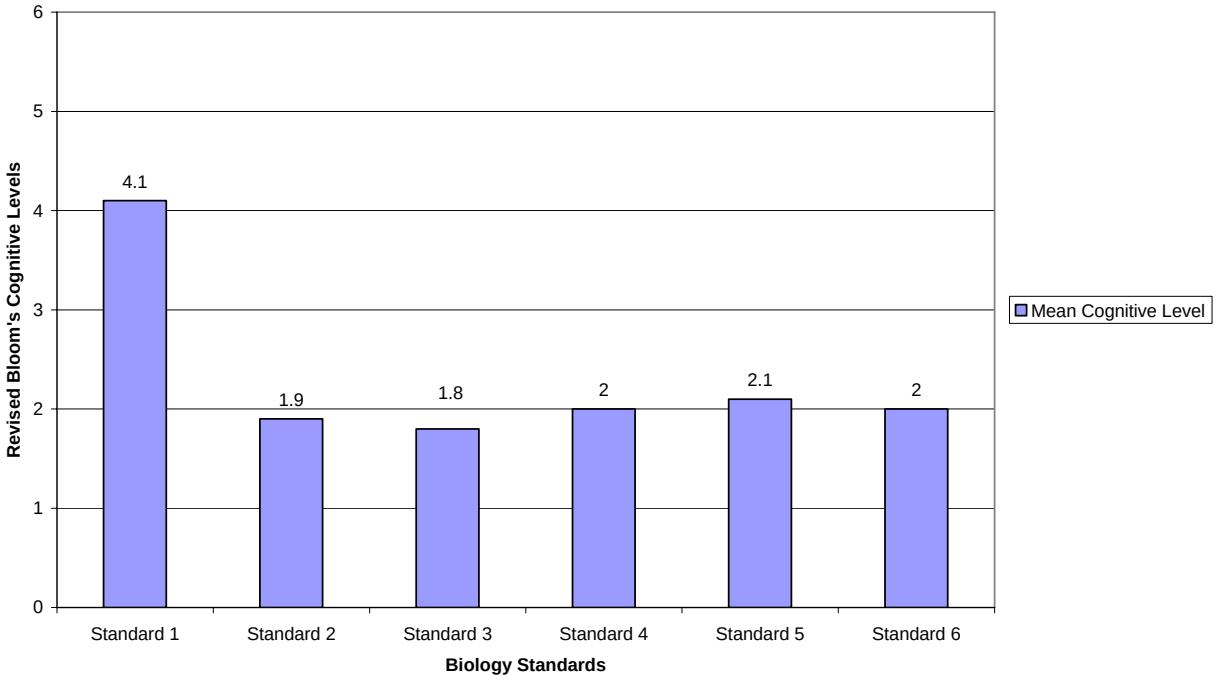
The cognitive processes specified in the SC Biology academic standards document were compiled by EOC staff for each indicator in each standard. The specific cognitive processes in the indicators were coded into the six revised Bloom's taxonomic categories (1, Remember; 2, Understand; 3, Apply; 4, Analyze; 5, Evaluate; or 6, Create), assigning a number from 1 to 6 to each indicator based on the cognitive processing category for the indicator. If an indicator listed more than one cognitive process, the highest level in the indicator was retained for this analysis. For example, Biology Standard 1, Indicator 5 has two verbs, "organize' and "interpret":

> Organize and interpret the data from a controlled scientific investigation by using mathematics, graphs, models, and/or technology.

In the revised Bloom's taxonomy "organize" is at a higher cognitive processing level (4, Analyze) than "interpret" (2, Understand). The higher level, 4, Analyze, was assigned to this indicator for this analysis.

This process resulted in a computer file listing the indicators and their corresponding cognitive levels to be used for further analyses and for comparisons to the cognitive processes elicited in the Biology 1/Applied Biology 2 test items evaluated for alignment. Since the cognitive processes were assigned to a numeric scale, the data describing the cognitive processing levels in the academic standards can be aggregated for various analyses. For example, the numbers corresponding to the cognitive processing levels of the indicators can be averaged across the standards to get a "big picture" view of the overall levels of cognitive processing for each Biology standard, as illustrated in Figure 1.

Figure 1
Average Cognitive Levels of Biology Academic Standards
(1=Remember; 2=Understand; 3=Apply; 4=Analyze; 5=Evaluate; 6=Create)



The data displayed in Figure 1 suggest that the Biology academic standards 2 through 6 specify, on average, a cognitive level of "2, Understand." The cognitive levels for the Standard 1 indicators range from "2, Understand" to "6, Create," with an average level of "4, Analyze."

Information on the levels of cognitive processing elicited by the test items was collected during the alignment meeting in April 2010. The educators reviewing the test items for the standards and indicators assessed by the items were also asked to rate the cognitive processing level for each item using the scale from 1 to 6 corresponding to the revised Bloom's taxonomy categories of cognitive processes. The reviewers were given copies of the revised Bloom's terms listed in Table 2 for use during the meeting. The reviewers were also asked to identify the knowledge dimension assessed by each item based on the dimensions described in the revised Bloom's taxonomy; that information has been provided to the SCDE.

For each item, a reviewer's identification of the primary indicator assessed and the reviewer's rating of the cognitive process elicited by the item were recorded on the data collection sheets. Since an item may elicit more than one level of cognitive processing, reviewers were asked to record all the cognitive processes they thought were required by the item to answer it successfully. For example, an item which asks a student to compare the effects on body organs of infectious and non-infectious diseases requires that a student know what infectious and noninfectious diseases are (1, Remember) before the student can compare (2, Understand) their effects. If more than one cognitive process level was recorded by a reviewer, the highest level was retained for this analysis.

Each test item was reviewed by three reviewers, resulting in multiple ratings of cognitive levels for each item. Of the 493 items reviewed, all three reviewers identified the same cognitive levels for 53.1% of the items. All three reviewers identified the same cognitive level or an adjacent level (for example, two reviewers might have identified level 3 for an item and the third reviewer identified level 2 for that item) for 81.5% of the items. The majority of the reviewers were in agreement on the cognitive levels they assigned to each item.

The analysis for determining depth-of-knowledge consistency was carried out by compiling the reviewers' cognitive process ratings for all the items measuring a standard and then comparing the ratings to the cognitive process level specified in the standard. This analysis began by sorting all the items and their associated cognitive process ratings by standard and aggregating all of the ratings on all the items measuring that standard. The cognitive process rating for each item measuring the standard was coded as being at a lower level than the cognitive process specified in the standards document, at the same level, or at a higher level. The frequencies of item ratings which were below, the same as, or above the cognitive process specified in the standards document were compiled and compared to the criteria for acceptable depth-of-knowledge consistency.

A test meets the criterion for depth-of-knowledge consistency if at least 50% of the items elicit cognitive process levels at (or above) the level specified in the standards document (Webb, 1999; Webb, 2002; Webb, Horton, & O'Neal, 2002). Stated another way, if 50% or more of the items on a test can be answered successfully using lower cognitive process levels than specified in the standards, the test would not fully meet this criterion. The criterion for depth-of-knowledge of 50% of the items at or above the cognitive levels stated in the academic standards was established for the Webb alignment methodology as described below:

> "Fifty percent, a conservative cutoff point, is based on the assumption that a minimal passing score for any one standard of 60% or higher would require the student to successfully answer at least some items at or above the depth-of-knowledge level of the corresponding objectives. For example, assume an assessment included six items related to one standard and students were required to answer correctly four of those items to be judged proficient – i.e. 67% of the items. If three, 50% of the six items, were at or above the depth-of-knowledge level of the corresponding objectives, then for a student to achieve a proficient score would require the student to answer correctly at least one item at or above the depth-of-knowledge of one objective." (Webb, Horton, & O'Neal, 2002, p. 5)

Thus the value for meeting the depth-of-knowledge criterion was chosen to assure that students answer at least some items measuring at the cognitive processing level stated in the standards document. If at least 50% of the items are at or above the cognitive levels in the academic standards the test is said to have "met" the criterion for depth-of-knowledge. The Webb methodology also provides some leeway for this criterion in that if between 40% and 50% of the

test items are at or above the cognitive process levels in the standards the test has "weakly met" the depth-of-knowledge criterion (Webb, 1999; Webb, 2002; Webb, Horton, & O'Neal, 2002).

Findings – Depth-of Knowledge

The findings from the depth-of-knowledge analysis of the Biology 1/Applied Biology 2 items are summarized in Table 3. The criterion for depth-of-knowledge test alignment was applied to all items reviewed (including both items field tested in 2009 and items scheduled to be field tested in 2010) in this analysis.

Table 3
Depth-of-Knowledge Alignment Analysis
Biology 1/Applied Biology 2 Items Reviewed

Levels of Cognitive Processes of Biology 1/Applied Biology 2 Items Compared to Cognitive Processes of Academic Standard Indicators:			Total Items Receiving Ratings for Cognitive Processes – Number (%)*
Lower Level Than Indicator - Number (%)	Same Level as Indicator - Number (%)	Higher Level Than Indicator - Number (%)	
300 (48.0)	182 (29.1)	143 (22.8)	625 (99.9)

* An item may have been counted more than once if different judges assigned different standards or indicators to the item.
Note: Percentages may not add to 100% due to rounding.

The Biology 1/Applied Biology 2 field test items met the depth-of-knowledge criterion for alignment (e.g., 52% of the items were judged to be assessing at or above the cognitive process levels specified in the Biology academic standards).

The Biology 1/Applied Biology 2 end of course tests will be constructed by sampling items representing the Biology standards and indicators. If the items in the bank for each indicator and standard are assessing at too low or too high a cognitive processing level, it may be difficult or impossible to construct test forms assessing the standards at the level of cognitive processing required by the academic standards. The Webb criterion for depth-of-knowledge consistency is intended to assure that the majority of items on a test are not assessing levels of cognitive processing below that expected in the state's academic standards. If an excessive proportion of items in the item bank assess at too low a level, the majority of items selected to construct a test may also be at too low a level to accurately assess students' achievement of the state academic standards. This may lead to erroneous interpretations of the test score results: we may conclude from the results that students scoring at levels meeting the performance standards on the test have the skills expected for their grade level, when in fact the test is not measuring at the levels expected in the state academic standards. The Depth-of-Knowledge results suggest that the Biology 1/Applied Biology 2 tests constructed from the items reviewed can assess the Biology academic standards at an appropriate level.

Range-of-Knowledge Correspondence

Background and Methodology – Range-of-Knowledge Correspondence

The Range-of-Knowledge measure poses the question, "Do the test items cover the breadth of knowledge and skills represented by the indicators in a standard?" As defined by Webb,

> "The range-of-knowledge criterion is used to judge whether a comparable span of knowledge expected of students by a standard is the same as, or corresponds to, the span of knowledge that students need in order to correctly answer the assessment items/activities." (Webb, Horton, and O'Neal, 2002, p. 7)

The criterion for an acceptable correspondence between the breadth of knowledge in the standards and the breadth of knowledge assessed on a test is that at least 50% of the indicators in a standard must have at least one item assessing the indicator. This criterion is based on the assumption that student performance should be assessed based on at least half of the domain of knowledge in a standard.

While the criterion of 50% of the indicators having associated test items may be appropriate for a test, it is probably too generous a criterion for an item bank. Since items are sampled from the group of items assessing a standard when new test forms are created, it seems reasonable to expect that there will be items in the bank assessing every indicator in the standard so that the measurement of the standard is representative of the knowledge expectations for students expressed in that standard. However, a criterion for this measure for item banks has not been published, so the 50% criterion will be used for this study.

Findings – Range-of-Knowledge Correspondence

Based on the Webb criterion, the Biology 1/Applied Biology 2 field test items reviewed meet this criterion: there were items identified assessing every Biology indicator, with a range of 5 to 25 items identified assessing each indicator. Thus sufficient items are available to assess the breadth of Biology content and skills. A detailed table displaying the numbers of items assessing each indicator is available in Appendix C.

Balance-of-Representation

Background and Methodology – Balance-of-Representation

Balance-of-Representation poses the question, "Are the numbers of items assessing an indicator proportional to the number of indicators tested for a standard, assuring that one indicator does not receive more weight than the other indicators in the scores from the assessment?" Webb defines Balance of Representation as follows:

> "The balance-of-representation criterion is used to indicate the degree to which one objective is given more emphasis on the assessment than another." (Webb, Horton, and O'Neal, 2002, p. 9)

The Balance-of-Representation measure assumes that every indicator within a standard has equal importance, so the number of items from each indicator assessing a standard should be proportional to the number of indicators assessed for the standard. Thus if six indicators assessed for a standard are each tested by a single item, their proportions would be balanced in the assessment of the standard. If, on the other hand, those six indicators from a standard were tested with one item each from five of the indicators and five items from the sixth indicator, the

overall score for the standard would not provide an accurate picture of student proficiency in that standard because almost half the items tested were based on only one indicator. The student who performs well on the indicator tested with five items but poorly on the remaining items testing one indicator each may be designated by the total score as meeting expectations for the standard, but actually may have some serious weaknesses in his or her understanding of all components of the content covered by the standard.

The Balance-of-Representation measure seems appropriate for use with the evaluation of the Biology 1/Applied Biology 2 assessment because the state academic standards do not assign higher priorities or weightings to some indicators compared to others – all indicators appear in the standards documents to have the same importance within their standards. This assumption that all indicators have equal importance within a standard underlies the Webb measure, also. The criterion for acceptable Balance-of-Representation is based on the evaluation of a test form. However, the criterion can be calculated for an item bank and provides a useful measure of the balance of items needed in the item bank to construct future forms with sufficient items for each indicator to generate accurate evaluations of student performance in each standard tested.

The criterion for Balance-of-Representation is based on an index calculated for each of the standards assessed. The index is calculated only on data from indicators which have at least one item assessing it; if there are no items assessing an indicator that indicator is not used in the calculation of the index. The index is calculated by computing the difference between the proportion of indicators assessed for the standard and the proportion of items assessing each indicator. The index rages from 0 to 1, with a value of 1 indicating a perfect balance between the proportion of indicators assessed for a standard and the proportion of items assessing each indicator. If most of the items assess a single indicator and the remaining indicators in the standard are assessed by only one item each, the index has a value less than 0.5. Index values of 0.7 or higher indicate that the items are distributed relatively equally among all the indicators. Index values of 0.7 or higher indicate that the Balance-of-Representation criterion has been "Met." Index values between 0.6 and 0.7 indicate that the criterion has been "Weakly Met."

Finding – Balance-of-Representation

A table displaying the index for each standard is available in Appendix C. The summary results from the analysis of Balance-of-Representation are listed below in Table 4. The Balance of Representation criterion was "Met" for every one of the Biology standards. Since at least 5 items were identified assessing each indicator, the data suggest sufficient items are available to produce more than one form of the test designed to provide comprehensive and balanced measurements of student performance on the Biology standards.

Table 4
Balance of Representation Results
Biology 1/Applied Biology 2 End of Course Test Alignment Review

Balance of Representation Index and Judgment: Did the Item Bank Meet the Criterion?*					
1: Scientific Inquiry	2: Structure and Function of Cells and Their Organelles	3: Energy Flow Within and Between Living Systems	4: Molecular Basis of Heredity	5: Biological Evolution and the Diversity of Life	6: Interrelationships Among Organisms and Biotic and Abiotic Components of Their Environments
0.81 Met	0.83 Met	0.92 Met	0.84 Met	0.85 Met	0.89 Met

*Are the items relatively equally distributed among the indicators for a standard, or do some indicators have most of the items and others few items? (Criterion measured with an index ranging from 0 to 1, with values 0.7 and above meet the criterion, values 0.60 – 0.69 weakly meet the criterion, and values less than 0.6 fail to meet the criterion.)

Technical Review of 2009 Biology1/Applied Biology 2 End of Course Field Tests

The Biology 1/Applied Biology 2 Spring 2009 field test consisted of six forms (designated Form 4 – Form 9) consisting of 60 multiple choice items each. Fifteen multiple choice items were common to all six forms ("anchor items") to facilitate placing all the items on all forms onto the same score scale. Each field test form contained items representative of the Biology 1/Applied Biology 2 test blueprint. The testing contractors for the SC Department of Education (SCDE) scored the field test forms and calculated item and test statistics.

The Spring 2009 field test data were reviewed on April 21, 2010 by a technical panel composed of three school district directors of assessment and one university professor of educational measurement (see Appendix D for a listing of the technical panel members). The information provided by the SCDE for the technical panel included item statistics on the 285 items field tested in Spring 2009, scored student responses from the Spring 2009 field test administration, and a draft technical report based on 2008 pilot testing and 2009 field testing data. The draft technical report contained information on the generation of the test items and construction of field test forms, administering, scoring, and scaling the field tests, and standard-setting procedures carried out for the SCDE in July 2009.

The 2009 Biology 1/Applied Biology 2 Field test was administered to a total of 30,107 students in Spring 2009. The field test was challenging to the students tested, with an overall item difficulty value of 0.47 (on a scale of 0.00 to 1.00). While the test was difficult, the technical panel found that the items are within an acceptable range of difficulty for a criterion-referenced test. Based on initial estimates from field test data provided by the SC Department of Education, approximately 61.5% of students tested would be expected to meet the minimum performance standard on the test in 2009.

The findings of the technical panel are summarized below as Positives, Areas for Concern, and Recommendations.

Positives

The field test form reliabilities ranged from .83 to .88 (Table 5), which are within an acceptable range for field test forms. The reliability coefficients of three of the forms (Forms 6, 8, and 9) exceeded the minimal criterion of .85 for an operational form. The field tests contained some items which are being "tried out" to see if they are acceptable to use in future operational forms "as is" or if they need to be revised because of technical concerns such as extreme difficulty or poor discrimination. The reliabilities of the field test forms are adversely affected by field test items having unacceptable technical characteristics; the reliabilities of future operational forms are expected to be higher than those of the field test forms because the operational forms will not contain the field test items having technical flaws.

Table 5
Reliability Coefficients for Spring 2009
Biology 1/Applied Biology 2 Field Test Forms

2009 Field Test Form	**Reliability Coefficient (KR_{20})**
Form 4	0.83
Form 5	0.84
Form 6	0.86
Form 7	0.83
Form 8	0.87
Form 9	0.87

Although the distributions of the raw scores from the field test forms showed a small positive skewness, indicating that more students had low scores than high scores, there was no apparent floor or ceiling effect on the scores from any of the field test forms.

There were acceptable ranges of difference between the average item difficulties between students who scored in the top 20% of overall performance and students who scored in the bottom 20%. The difference in average item difficulty between these groups across the six field test forms was 0.41 (0.69 average item difficulty for students in top 20% of score distribution compared to 0.28 average item difficulty for students in lowest 20% of distribution). The difference observed between the performances of the two groups suggests that the field test forms are adequately discriminating between high ability and low ability students.

There was a reasonable separation between the median raw scores of students on each form and the scores expected by chance alone, indicating that, while the field tests were challenging, students are performing well above the chance level (Table 6).

Table 6
Comparison of Median Raw Scores to Raw Scores
Based on Chance Responses Alone
Spring 2009 Biology 1/Applied Biology 2 Field Test

2009 Field Test Form	Median Raw Score	Raw Score Based on Chance Alone
Form 4	26	15
Form 5	27	15
Form 6	29	15
Form 7	28	15
Form 8	29	15
Form 9	29	15

There were no items on any of the field test forms which were identified as exhibiting significant differential item performance (DIF). The finding of a significant DIF statistic for an item is a flag that the item should be examined carefully to rule out the possibility that successful response to the item is the result of item characteristics which make it easier for one demographic group to respond correctly for reasons other than knowledge or skill in the subject area tested.

Areas for Concern

Some items had technical values outside the acceptable guidelines for the statistics:

- 41 of the 285 items (14.3%) were flagged because their item difficulty value was less than 0.3 indicating the item was extremely difficult, suggesting that students could be guessing the answer because they did not know or understand the concept being tested (perhaps because they had not been taught the concept), or that the item was faultily constructed;
- 88 of the 285 items (30.9%) had low (less than 0.2) or negative item discrimination statistics indicating the item did not properly discriminate between students with high overall ability in the subject from students with low overall ability; the low item discrimination statistics could be related to the extreme difficulty of some items (33 of the items flagged for item discrimination were also flagged for extreme difficulty), or to aspects of the wording or format of the items.

Students taking the 2009 Biology 1/Applied Biology 2 end of course field test were asked to record their starting and ending time for the test administration on their answer sheets. The length of time reported by the majority of students to complete the test was 2 hours or greater across the nine test forms (from 25% to 95% of the students reported taking 2 hours or more, with approximately 90% of the students reporting 2 hours or more on six of the nine forms). If the test takes similar lengths of times for administration in the future, this may lead to problems with scheduling the tests in high schools.

The field test sample, while representative of the student population currently enrolled in Biology 1/Applied Biology 2, did not include all high school students in the cohort grade level. Approximately 5% of the students in the field test samples were enrolled in Applied Biology 2, with the remaining students enrolled in Biology 1. The Biology 1 course is typically a year long course taken by students in an academic curriculum, while Applied Biology 2 is a two-year course for students pursuing an applied curriculum. Not all high school students are currently

required to take a Biology course, so the number of students enrolled in Applied Biology 2 included in the field test sample may not represent the increased number of students taking Applied Biology 2 expected when the Biology 1/Applied Biology 2 assessment is required to meet federal ESEA guidelines and thus must be taken by all high school students. In general, the end of course test performance of students enrolled in the applied high school courses has been lower than that of students enrolled in the non-applied courses. This trend is evident in the field test data for Biology 1/Applied Biology 2: information from the SCDE from the standard setting analysis indicates that approximately 30% of the students taking the Applied Biology 2 field test in 2009 were projected to meet the minimum performance standard (scale score of 70 or higher), while approximately 63% of the students taking the Biology 1 field test were projected to score 70 or higher.

Summary and Recommendations from the Alignment and Technical Reviews

The results from the review of the alignment of the Biology 1/Applied Biology 2 end of course test items reveal that they are well aligned to the state academic course standards. The field test items analyzed met all four of the Webb (2002, 1999) criteria for test alignment:

5. Categorical Concurrence: there were more than 6 items assessing each standard and the proportions of items assessing the standards were similar to the proportions of items for each standard listed in the test Blueprint;
6. Depth-of-Knowledge: 52% of the items assessed at or above the cognitive process levels specified for the academic standard indicators;
7. Range-of-Knowledge: there were from 5 to 25 items identified for assessing each indicator, suggesting that sufficient items were available for each indicator to construct more than one test form which would meet the specifications of the test Blueprint;
8. Balance-of-Representation: the indexes of balanced representation for each standard met the criteria for tests which assess the content in a comprehensive and balanced manner, without under- or over-representing some objectives compared to others.

The technical review concluded that, while the field test was difficult, the results were found to be within an acceptable range for a field test. The test reliabilities were also within an acceptable range for field tests and the item statistics for most of the items were also within acceptable ranges. However, some items were flagged for being exceptionally difficult and some items were flagged for poor discrimination between students having overall high scores on the rest of the items on the field test and students having overall low scores. There were indications from the student survey responses that the test may have taken two hours or more for administration, which may cause conflicts in high school class schedules. Finally, the population of students taking the Biology 1/Applied Biology 2 end of course test is expected to change when the course becomes required for all students such that more students enrolled in Applied Biology 2 will be taking the test than were evident in the field test sample.

Recommendations

1. The items flagged for out-of-range item statistics should be reviewed to identify the reasons for the observed item statistics and, if necessary, revised before use in future test forms.

2. The Biology 1/Applied Biology 2 end of course test is not a “timed” test: students can work on the test as long as needed to complete it. The length of time needed to administer the test

which was reported from the field test survey (90% of the students reporting 2 hours or more on six of the nine forms) should be reviewed and if verified, the SCDE should consult with school district personnel regarding ways to address the scheduling needed to provide appropriate time for administering the test.

3. Since the student sample participating in the Biology 1/Applied Biology 2 field test may not be fully representative of the student cohorts required to take the assessment in the future because of federal requirements, the scaling and standard setting of the Biology 1/Applied Biology 2 test should be revisited when the test becomes required and more representative information becomes available from the broader population tested. The anticipated increase in the number of students enrolled in Applied Biology 2 when the test becomes required may result in lower overall test performance and may change the distribution of scores enough to warrant review and possible adjustment of the scaling and performance standards.

References Cited

American Educational Research Association (2003). “Standards and Tests: Keeping Them Aligned,” *Research Points: Essential Information for Education Policy*, Vol. 1, No. 1, Spring 2003.

Anderson, L. W., & Krathwohl, D. R. (2001). A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom’s Educational Objectives. Boston: Allyn and Bacon.

Council of Chief State School Officers (2002). *Models for Alignment Analysis and Assistance to States,* Washington, DC: Council of Chief State School Officers.

South Carolina Department of Education (December 2009). *Spring 2009 Field Test Technical Report for Biology 1/Applied Biology 2 (DRAFT)*, Columbia, SC.

South Carolina Department of Education (2008). *South Carolina Academic Standards for English Language Arts*, Columbia, SC.

South Carolina Department of Education (November 2005). *South Carolina Science Academic Standards*, Columbia, SC.

Webb, N. L. (1999). *Alignment of Science and Mathematics Standards and Assessments in Four States.* (Research Monograph No. 18) Madison, WI: University of Wisconsin-Madison, National Institute for Science Education.

Webb, N. L. (April 2002). “An Analysis of the Alignment Between Mathematics Standards and Assessments for Three States.” Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Webb, N. L., Horton, M., & O’Neal, S. (April 2002). “An Analysis of the Alignment Between Language Arts Standards and Assessments for Four States.” Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

APPENDIX A

Test Blueprint for Biology 1/Applied Biology 2 End of Course Test

Test Blueprint for
Biology
OPERATIONAL FORMS

Item Allocation Rules for Test Form Construction	Number of Items
Standard B-1: The student will demonstrate an understanding of how scientific inquiry and technological design, including mathematical analysis, can be used appropriately to pose questions, seek answers, and develop solutions.	10-11
Standard B-2: The student will demonstrate an understanding of the structure and function of cells and their organelles.	9-10
Standard B-3: The student will demonstrate an understanding of the flow of energy within and between living systems.	9-11
Standard B-4: The student will demonstrate a understanding of the molecular basis of heredity.	10-12
Standard B-5: The student will demonstrate an understanding of biological evolution and the diversity of life.	9-11
Standard B-6: The student will demonstrate an understanding of the interrelationships among organisms and the biotic and abiotic components of their environments.	9-11

APPENDIX B

Biology 1/Applied Biology 2 Alignment Review Panel Members

Biology 1/Appilied Biology 2 End-of-Course Test
Alignment Review Panel
April 22, 2010
SC State Library, Columbia

Name	District
Terri Dees	Spartanburg 7
Murray Eicher	Charleston
Rona Ellis	Orangeburg 5
Amy Hawkins	Anderson 5
Rebecca Jackson	Dorchester 2
Laura Lawrence	Abbeville
Trey McGarity	Chester
Henry Mack	Laurens 56
Suwarna Patel	Marion 1
Deborah Pogue	Horry
Nicole Riddle	Richland 1
John Sacco	York 4-Ft Mill

APPENDIX C

Table of Alignment to Standards and Indicators Results

Table C-1
Biology 1/Applied Biology 2 End of Course Test Alignment to Standards and Indicators

Biology Standard & Indicator	No. Items Identified	Total Items Identified for Standard*	Proportion of Items in Standard	Balance of Representation Index	Met Balance Criterion?	Met Range of Knowledge Criterion?
1.1	12		0.10084			
1.2	15		0.12605			
1.3	8		0.067227			
1.4	19		0.159664			
1.5	23		0.193277			
1.6	18		0.151261			
1.7	11		0.092437			
1.8	5		0.042017			
1.9	8	119	0.067227	0.814192	Yes	Yes
2.1	5		0.053191			
2.2	13		0.138298			
2.3	8		0.085106			
2.4	13		0.138298			
2.5	17		0.180851			
2.6	18		0.191489			
2.7	14		0.148936			
2.8	6	94	0.06383	0.827128	Yes	Yes
3.1	13		0.142857			
3.2	18		0.197802			
3.3	16		0.175824			
3.4	14		0.153846			
3.5	19		0.208791			
3.6	11	91	0.120879	0.917582	Yes	Yes
4.1	12		0.096			
4.2	7		0.056			
4.3	15		0.12			
4.4	15		0.12			
4.5	15		0.12			
4.6	25		0.20			
4.7	19		0.152			
4.8	7		0.056			
4.9	10	125	0.08	0.843556	Yes	Yes
5.1	7		0.070707			
5.2	12		0.121212			
5.3	15		0.151515			

Biology Standard & Indicator	No. Items Identified	Total Items Identified for Standard*	Proportion of Items in Standard	Balance of Representation Index	Met Balance Criterion?	Met Range of Knowledge Criterion?
5.4	16		0.161616			
5.5	25		0.252525			
5.6	15		0.151515			
5.7	9	99	0.090909	0.854257	Yes	Yes
6.1	21		0.216495			
6.2	14		0.14433			
6.3	8		0.082474			
6.4	21		0.216495			
6.5	16		0.164948			
6.6	17	9	0.175258	0.891753	Yes	Yes

* An item may have been counted more than once if different judges assigned different standards or indicators to the item.

APPENDIX D

Biology 1/Applied Biology 2 Technical Review Panel

Biology 1/Applied Biology 2 Technical Review Panel
April 21, 2010
Columbia, SC

Dr. Christine DiStefano
Associate Professor
Educational Research
Department of Education Studies
College of Education
University of South Carolina

Dr. Kevin Andrews
Research Specialist
Instruction & Accountability
Rock Hill School District 3

Dr. Janet S. Rose-Baele
Executive Director
Assessment & Accountability
Charleston County School District

Mrs. Missy Wall-Mitchell
Director of Accountability
School District 5 of Lexington & Richland Counties

SC Department of Education Staff
Mr. Joe Saunders
Education Associate
Office of Assessment

Mr. David Potter
Director of Research