South Carolina
Alternate Assessment
(SC-Alt)

South Carolina's Alternate Assessment, SC-Alt
Spring 2010 Operational Administration

Technical Report
October 29, 2010

American Institutes for Research
South Carolina Department of Education

TABLE OF CONTENTS

Introduction

This report details the design, development, and spring 2010 operational test results for the South Carolina Alternate Assessment (SC-Alt). The SC-Alt consists of four content areas: English language arts (ELA), mathematics, science, and social studies. The assessments are administered across three grade-bands: 3–5, 6–8, and 10. The 2010 field-testing of the new high school biology assessment is reported in a separate chapter of this technical report.

Chapter 1: *Development of Alternate Assessment in South Carolina* describes the background of the alternate assessments in South Carolina, the format of the previous assessments, and the need for a new alternate assessment.

Chapter 2: *Test Development* describes the design of the alternate assessment and the development of tasks and items to measure academic growth among students who have significant cognitive disabilities. The Student Placement Questionnaire (SPQ), a unique feature designed to maximize the efficiency of teacher and student testing time, is described and thoroughly reviewed. The development of a vertical scale linking grade-appropriate tasks across grade levels and complexity levels within grades is described.

Chapter 3: *Spring 2010 Operational Test Administration* details the spring 2010 operational test administration in ELA, mathematics, science, and social studies; test administrator training; use of the SPQ; measures taken to ensure the accuracy of scoring; and the maintenance of test security.

Chapter 4: *Setting Performance Standards* describes the procedures for setting performance standards. The chapter includes a description of the Item Descriptor (ID) Matching procedure, the goals of the standard-setting workshops, the composition of the standard-setting panels, the workshop activities, and the panels' recommended performance standards.

Chapter 5: *Technical Characteristics and Interpretation of Student Scores* reviews technical topics including analysis and scaling, reliability of test scores, the procedures used to calculate internal consistency reliability estimates, and classification accuracy estimates.

Chapter 6: *Score Reports* describes the score reporting system for SC-Alt with emphasis on the Individual Student (Family) Report (see Appendix H) from which the summary reports are derived, the information contained in the various reports, and their intended uses.

Chapter 7: *Student Performance Data from the Spring 2010 Administration* provides an overview of statewide achievement on the SC-Alt, based on the spring 2010 operational test administration.

Chapter 8: *Validity* reports on content validity and convergent and discriminant validity topics as well as the validity of the SPQ and the operational performance of the tailored assessment under the SPQ's start/stop rules.

Chapter 9: *High School Biology Field Test* describes the forms design and sampling for the spring 2010 high school biology field test. Both planned and realized sample statistics are reported, followed by a summary of the biology item data review and standard-setting workshop results.

Chapter 1: Development of Alternate Assessment in South Carolina

Overview of the State Assessment System

The South Carolina Assessment System includes the South Carolina Palmetto Assessment of State Standards (PASS), the High School Assessment Program (HSAP), and the End-of-Course Examination Program (EOCEP). These state-level assessments are required by the Education Accountability Act of 1998 (EAA) as amended May 2008 and are aligned with the state's academic standards for each subject and grade level.

- PASS measures the performance of all public school students in grades 3 through 8 in the content areas of English language arts (ELA), mathematics, science, and social studies.
- HSAP measures the performance of high school students in ELA and mathematics and is used both as one criterion for eligibility to receive a high school diploma and as the primary source for reporting the federally mandated data required by the No Child Left Behind Act (NCLB).
- EOCEP is administered in gateway courses at the high school level. The physical science EOCEP examination is counted for participation purposes for NCLB reporting.

The EAA establishes a performance-based accountability system that includes all students. This act supports South Carolina's commitment to public education and a conviction that high expectations for all students are a vital component of improving academic education.

The goals of the state assessment system are as follows:

- Increasing academic performance of all children and, ultimately, raising high school graduation rates
- Implementing rigorous academic achievement standards that are aligned with the South Carolina curriculum standards
- Improving instruction based, in part, on the implementation of these higher standards
- Using the results of challenging assessments that measure student performance relative to these standards

Another goal is to inform various audiences—teachers, school administrators, district administrators, South Carolina State Department of Education (SCDE) staff, parents, and the public—of the status of academic performance and of the progress of public school students toward meeting South Carolina's academic achievement standards.

The South Carolina academic standards form the basis for alignment across the state education system for district and school curricula, classroom instruction, units of study, and learning experiences. **The academic standards are the basis for all assessments in the state assessment system, including alternate assessment.**

Purpose of the South Carolina Alternate Assessment

The purpose of the alternate assessment based on alternate achievement standards is to capture and evaluate the performance of students who have traditionally been excluded from statewide testing programs and to improve instruction for these students by promoting appropriately high expectations and the inclusion of these students in state accountability for district report cards and for adequate yearly progress (AYP) reporting at the school, district, and state levels.

Description of the South Carolina Alternate Assessment

The South Carolina Alternate Assessment (SC-Alt) is administered to students who have been determined by the Individualized Education Program (IEP) team to be unable to participate in the general state assessments even with appropriate accommodations. It is an alternate assessment on alternate achievement standards to the PASS for students in grades 3–8 and the HSAP and Physical Science EOCEP for high school students.

The test is administered to students who meet the participation criteria for alternate assessment and who are of the ages of typical students in grades 3–8 and 10. Students who are ages 8–13 (the typical ages for grades 3–8) are assessed in ELA, mathematics, science, and social studies. Students who are 15 (the typical age of students in grade 10) are assessed in ELA, mathematics, and physical science.

The SC-Alt consists of a series of performance tasks that are scored by the test administrator (teacher) as they are administered. The performance tasks are scripted activities, and each task contains four to eight related items. The items have a scaffolded scoring script to reduce the complexity of the item when students do not respond successfully on the first attempt. All items are linked to the South Carolina academic content standards through the South Carolina Alternate Assessment Standards and Measurement Guidelines (ASMGs) and South Carolina Alternate Assessment Extended Standards. The ASMGs and Extended Standards are linked explicitly to the South Carolina academic standards for grades 3–8 and 10, although at less-complex or prerequisite levels. Currently there are ASMGS for social studies, but in the areas of ELA, mathematics, science and biology, the ASMGs have been replaced by Extended Standards linked to the revised ELA (2008), mathematics (2007), science and biology (2005) standards.

The SC-Alt has three forms: elementary, middle, and high school. Students' assignment to forms is based on their age on September 1 of the tested year; 8- to 10-year-olds take the elementary form, 11- 13-year-olds take the middle school form, and 15-year-olds take the high school form.

The assessment is designed to minimize teacher and student testing burden by administering only those items that are well-suited to a student's achievement level. The test administrator completes a Student Placement Questionnaire (SPQ) to determine the most appropriate starting task for the student. Tasks are arranged in ascending order of difficulty. Once the appropriate starting task is identified, test administrators continue to administer tasks until the student can no longer respond successfully.

The first operational administration of the SC-Alt was conducted during a seven-week testing window during spring 2007 in ELA, mathematics, and science. A census field test was conducted during the same assessment window for the social studies assessment. In spring 2009,

an embedded field test was added. Documentation related to the 2010 operational administration is the focus of this Technical Report.

Background on Alternate Assessment Development in South Carolina

The 1997 amendments to the Individuals with Disabilities Education Act (IDEA '97) created the mandate to include all children, including children with significant disabilities, in state testing and accountability systems. The vision for the South Carolina alternate assessment system was initiated in early 1998 in response to the IDEA '97 regulations. This vision has driven the development and revision of alternate assessment in South Carolina.

A core team of staff from the SCDE Offices of Exceptional Children, Assessment, Research, and Curriculum and Standards met in March 1998 to develop a plan for designing an alternate assessment to meet the IDEA mandate and to be included in the state assessment system. The team's first steps were to convene a steering committee and seek technical assistance from the Mid-South Regional Resource Center (MSRRC) to explore strategies for designing an alternate assessment.

The Alternate Assessment Steering Committee was convened on May 12, 1998, to assist SCDE in determining how to include students with significant cognitive disabilities in statewide assessments. The committee comprised parents, special education and general education teachers, administrators, and representatives from other agencies. Dr. Ken Olsen of MSRRC provided the committee with technical assistance, including information on IDEA requirements, examples of options that some states were using or considering, and research available on alternate assessment. He facilitated a process that allowed the Steering Committee to reach shared foundational beliefs, address eligibility criteria and content and performance standards, and develop plans.

To ensure that all students, including students with significant disabilities, are included in the testing and accountability systems and have appropriate access to instruction in the South Carolina academic standards, the Steering Committee determined that the alternate assessment would be based on the following principles:

- All children can learn, be expected to meet, and be challenged to meet high standards.
- Special education is an extension and adaptation of the general education program and curriculum, rather than an alternate or separate system.
- The South Carolina State Board-approved standards are the foundation for all students, including students with unique needs and abilities.
- Measurement and reporting must be defensible in terms of feasibility, validity, reliability, and comparability.
- Results of the state standards-based program must be used to improve planning, instruction, and learning.
- An alternate assessment is appropriate for the few students for whom the state assessment, even with accommodations, is not appropriate.
- The alternate assessment is designed for a diverse group of students and should be flexible enough to address their individual needs.

The committee articulated these goals for the alternate assessment:

- To provide evidence that students have acquired the skills and knowledge necessary to become as independent as possible
- To document the student's performance and the performance of the programs serving the student
- To merge instructional best practice, instruction in state standards, and assessment activities
- To provide information in the development of curriculum that is responsive to the student's needs

The Steering Committee created the following participation guidelines to guide IEP team decisions regarding students who should participate in the alternate assessment:

- The student demonstrates significant cognitive disabilities and adaptive skills, which result in performance that is substantially below grade-level achievement expectations even with the use of accommodations and modifications.
- The student accesses the state-approved curriculum standards at less complex levels and with extensively modified instruction.
- The student has current adaptive skills requiring extensive direct instruction and practice in multiple settings to accomplish the application and transfer of skills necessary for application in school, work, home, and community environments.
- The student is unable to apply or use academic skills across natural settings when instructed solely or primarily through classroom instruction.
- The student's inability to achieve the state grade-level achievement expectations is not the result of excessive or extended absences or social, cultural, or economic differences.

NOTE: The term significant cognitive disabilities was added by the South Carolina Alternate Assessment Advisory Committee to the criteria after the passage of the NCLB December 2003 regulations on alternate assessment.

The Steering Committee recommended that the state develop a portfolio collection of evidence of student progress toward the South Carolina academic standards similar in design to the Kentucky Portfolio Alternate Assessment. The committee also recommended that SCDE prepare a Request for Proposal (RFP) for a contractor to develop the alternate assessment. Advanced Systems in Measurement and Evaluation Inc. (ASME), which later became Measured Progress, was awarded the contract. This company, along with the Inclusive Large Scale Standards and Assessment (ILSSA) project at the University of Kentucky, began work with SCDE on the design of PACT-Alt.

A work group was convened to define the domain for instruction and assessment. To ensure that the South Carolina curriculum standards were the foundation for all students, including students with unique needs and abilities, the work group developed adaptations of the curriculum standards. The work group comprised special education teachers, regular education teachers,

parents, administrators, higher education personnel, representatives from community agencies, and SCDE personnel. The work group process, which was facilitated by staff from MSRRC, focused on the prerequisite skills found primarily in the curriculum standards in prekindergarten through grade 2.

The work group affirmed that special education services must operate as an extension of the general education program and curriculum rather than as an alternate or separate system. The standards in this initial document were identified as concepts that every student, including students with moderate to severe disabilities, should know or be able to perform. These selected standards, which focused on skills that were deemed essential and attainable for every student, were directed toward the following goals:

- Enhancing the quality of students' communication skills
- Improving the quality of students' everyday living
- Improving students' ability to function in society and promote in them an acceptance of and respect for self and others
- Preparing students for transition into adult living
- Moving students toward independence, which may range from a level of self-care with assistance to total self-sufficiency

The extensions were based on the state academic content standards in prekindergarten through grade 2. For each selected standard, examples of essential real-world performance skills were developed. The articulation of these performance skills was designed to provide the rationale for teaching the standards and to serve as guides for teachers and parents regarding how the student demonstrated a skill. The committee specified that these performance skills could be accomplished in home, school, and community environments through a variety of individualized communication systems and might incorporate a variety of supports, such as physical assistance, physical prompts, verbal prompts, and technology. The document, *The Extensions and Adaptations of the South Carolina Curriculum Standards for Students Participating in Alternate Assessment*, became the focus of the portfolio assessment process, HSAP-Alt performance tasks, and professional development training. In 2002, this document was revised and renamed the *Resource Guide to the South Carolina Curriculum Standards for Students in Alternate Assessment*, but it was still aligned to curriculum standards for prekindergarten through grade 2. This work was based on the IDEA requirements and the thinking at the time about how students with significant cognitive disabilities should be included in the general education curriculum and assessment.

Beginning with the 2000–01 school year, students in grades 3–8 who met the participation criteria for alternate assessment were assessed with the portfolio assessment, PACT-Alt. In 2003, a high school assessment, HSAP, which was designed to meet AYP requirements, was added to the state assessment system, and an alternate to HSAP was developed to measure student proficiency in ELA and mathematics. A Stakeholder Committee with expertise in high school instruction of students with significant cognitive disabilities and academic standards was convened to guide the development of the high school alternate assessment, HSAP-Alt. The committee recommended designing an assessment based on performance on a series of tasks

linked to the state curriculum standards. The HSAP-Alt consisted of a series of scripted performance tasks in ELA and mathematics with scaffolded administration and scoring procedures aligned with the *Resource Guide to the South Carolina Curriculum Standards for Students in Alternate Assessment*.

One critical piece of the development and implementation process of PACT-Alt and HSAP-Alt was the provision of intensive professional development related to standards-based instruction, much of it based on the work of Harold Kleinert and Jacqui Farmer Kearns. A resource for professional development was their book, *Alternate Assessment: Measuring Outcomes and Supports for Students with Disabilities*. Professional development was essential to the implementation of the portfolio assessment because the teacher was responsible for teaching a student the content related to the academic standards, assessing the student's progress, and providing evidence of the instruction and progress in the portfolio. Prior to the implementation of the alternate assessment and the IDEA requirement to include students with disabilities in the general education curriculum, many students with disabilities, especially those with significant disabilities, and their teachers had been excluded from standards-based instruction and professional development related to academic standards.

Transition from PACT-Alt and HSAP-Alt to SC-Alt

After seeking input on the vision of a new alternate assessment on alternate achievement standards from the Advisory Committee and teachers who were conducting alternate assessment, SCDE wrote an RFP for the redesign or design of the alternate assessment system. The design was to be consistent with South Carolina's commitment to the instruction and assessment of students with significant cognitive disabilities and NCLB requirements. The focus was to be on grade-level academic standards. The new system was to address concerns related to teacher burden and time involved in assessment while supporting improved instruction based on state academic achievement standards. Extensive training for test administrators was to be integrated into the design of the assessment.

In September 2004, a contract was awarded to the American Institutes for Research (AIR) to assist the state in revising the alternate assessment. AIR managed the administration and analyses of the PACT-Alt and HSAP-Alt assessments during the 2004–05 and 2005–06 school years while developing the new alternate assessment, the South Carolina Alternate Assessment (SC-Alt), with SCDE.

American Institutes for Research

The American Institutes for Research (AIR) has more than 50 years of experience as a nonprofit organization dedicated to assessment, behavioral science, and educational research. AIR developed the South Carolina HSAP and the EOCEP programs and has enjoyed a successful collaboration with SCDE for a number of years.

Chapter 2: Test Development

The South Carolina academic content standards are the basis for alignment across the state for district and school curricula, classroom instruction, units of study, and learning experiences. The curriculum standards are the basis for the Palmetto Assessment of State Standards (PASS), the High School Assessment Program (HSAP), the End-of -Course Examination Program (EOCEP), and the alternate assessment. An initial step in the design of the new assessment was developing Assessment Standards and Measurement Guidelines (ASMGs).

Development of the Assessment Standards and Measurement Guidelines

In April 2005, a committee comprising South Carolina special education teachers, content specialists, SCDE staff, and AIR staff designed the ASMG document to support the new assessment development. The process involved extending the state academic standards in ELA, mathematics, science, and social studies in grade-bands 3–5, 6–8, and 10 to be accessible to students with significant cognitive disabilities. This document replaced the *Resource Guide to the South Carolina Curriculum Standards for Students in Alternate Assessment.*

The ASMGs are the foundation for the development of the assessment tasks for the SC-Alt. The ASMGs in each content area are distillations of the essence of South Carolina curriculum standards at each grade level.

Each content area committee reviewed the large array of standards and prioritized those in grade-bands 3–5, 6–8, and 10 that they deemed most important to students now and in the future. They then reduced the complexity of these standards, while retaining the essence of the grade-level content knowledge and skills, to make the academic standards appropriate and accessible for students with significant cognitive disabilities. The committee was careful to address both the depth and the breadth of the academic standards and used professional judgment based on experience with the population and the content to determine the standards to be assessed. The resulting document provides the link to the grade-level standards and indicators in the state academic standards. The measurement guidelines give task writers and teachers the specificity necessary to translate the assessment standards into assessment tasks and items and classroom instruction. A list of individuals who were involved in this process is included in each ASMG content document.

NOTE: The ELA committee recommended that the standards in the Research Goal not be included in the assessment standards. The rationale for this recommendation was that this goal is not tested to any great extent in PACT because this content is primarily taught and assessed at the classroom level. Committee members, however, indicated that the Communication Goal included standards that they deemed very important to this population, and they recommended including assessment standards for this strand.

The State Board of Education adopted mathematics and ELA academic standards in August 2007 and May 2008. State Board of Education requirements call for replacing the high school physical science end-of-course assessment for all students with a biology end-of-course assessment. The adoption of these new standards, which occurred outside the cyclical review timetable, and the replacement of the physical science end-of-course assessment with the biology end-of-course assessment have a direct impact on the ongoing schedule for developing additional tasks for the task pool.

During the 2007 and 2008 school years, committees of special educators and general educators met to extend the revised ELA, mathematics, and science academic standards and the biology end-of-course standards. These documents were designed to provide specificity for instruction as well as assessment, so the committees extended all standards and indicators including those for non-tested grades. These documents, referred to as the Extended Standards, replace the ASMGs in ELA, mathematics, and science. The Extended Standards provide extensions for all grade levels including those that are not tested and guidance to assist educators with instructional access to the state academic standards.

Stakeholder Input into the Development of the SC-Alt

To ensure the validity of the overall assessment process, a great deal of time and effort was spent obtaining input from various sources, including the State Alternate Assessment Advisory Committee, classroom teachers, parents, and other agency personnel.

South Carolina State Alternate Assessment Advisory Committee

The State Alternate Assessment Advisory Committee meets to provide oversight to the SC-Alt. The committee includes members of the original Alternate Assessment Steering Committee and the High School Stakeholder Committee. The committee also includes parents, special educators, representatives of higher education, content specialists, special education directors, and district test coordinators. Additional members include representatives from the Department of Disabilities and Special Needs, the University of South Carolina School of Medicine, the South Carolina Assistive Technology Project, the South Carolina Interagency Deaf-Blind Project, the Autism Society of South Carolina, and Pro-Parents of South Carolina.

The Advisory Committee provided input on its expectations for the revised alternate assessment during the first meeting with the contractor, AIR, on November 5, 2004. SCDE and AIR staff reported each step of the development process to the Advisory Committee at each meeting and sought its advice and recommendations.

Early Development Activities

At the recommendation of the Advisory Committee, AIR item writers visited classrooms in South Carolina during January and February 2005 to observe teaching strategies and materials that were in use. They also reviewed PACT-Alt portfolios for examples of evidence that teachers used to demonstrate progress toward proficiency on grade-level standards and examined the characteristics of the HSAP-Alt performance event in order to build on the existing system.

Teacher focus groups convened during January 2005 obtained feedback from teachers on the types of tasks they believed were appropriate, the protocol format they preferred, and the materials they recommended for inclusion in the assessment.

Qualified item writers employed by AIR were trained to write tasks and items specifically aligned with the ASMGs. Item writing teams included AIR staff with expertise in the content areas; alternate assessment specialists; and consultants in the areas of instruction of students who are blind and visually impaired, students who are deaf and hard of hearing, and students with cognitive disabilities.

On February 14, 2006, prior to the development of science and social studies tasks, SCDE staff and the AIR alternate assessment specialist provided additional training to the writing teams. The training was based on *Designing from the Ground Floor*, materials developed by the National Alternate Assessment Center (2005).

Consideration of universal design was a focus throughout the development process. Items including passages and response options were developed to use objects, pictures, picture symbols, words, and numbers. Several tasks in all four content areas and at different levels of complexity were piloted with South Carolina teachers and students in March and May 2005. AIR staff then interviewed the pilot teachers to determine the item characteristics and parameters that teachers believed worked well or did not work.

Summary of the Development and Review of the Original SC-Alt Tasks

- The task and item development process began with the creation of task kernels. AIR was primarily responsible for the majority of task kernels, with input from SCDE and teachers in South Carolina. Task kernels are basic ideas for an assessment activity, stimulus materials, and purpose, which, based on their relation to the South Carolina ASMGs, were used to develop a task and its items.
- SCDE reviewed the task kernels and provided feedback to AIR on which kernels were acceptable, which were unacceptable, and which needed revision. These reviews included alignment with the ASMGs.
- AIR item writers developed the items and stimulus materials. These items were reviewed internally by the content experts for clarity, quality, and alignment with the ASMGs.
- Following the comprehensive AIR internal review, the tasks and items underwent technical review by AIR to ensure that the items were properly keyed and scaffolded, the instructions were appropriate, the stimulus materials were interpretable, and the items were generally consistent in design with other tasks and items under development.
- Items that passed internal review by the AIR development staff were reviewed by the senior content lead for each content area and the senior alternate assessment specialist. This review ensured that within the content area, tasks and items followed the design of the assessment and were consistent with respect to format, presentation, and general administration procedures.
- Before items were passed to SCDE, the project director reviewed all items to ensure that they were consistent with the foregoing factors across content areas and grade-bands.
- Following the final internal AIR review, items were passed to SCDE for its review. During this process, SCDE staff, including content specialists, special educators, and assessment specialists, provided feedback to AIR on the design of the tasks and items, the alignment of items to the ASMGs, and the appropriateness of the items for use in South Carolina. Some items were revised by SCDE to improve alignment with the ASMGs.
- Approved items were placed into tasks for a small-scale tryout, conducted by AIR with the assistance of teachers in South Carolina and Northern Virginia and AIR staff. These tryouts provided invaluable information regarding the clarity of instructions, the utility of

the stimulus materials, and the success of the items and tasks in producing expected responses. Items that showed obvious problems were revised or discarded.

- After changes were made to the prototypes as a result of the pilots and tryouts, a committee of South Carolina teachers was convened on July 12, 2005, to review the revised tasks and provide further input and recommendations.

Content, Bias, and Sensitivity Reviews

Once small-scale tryouts were concluded, AIR, SCDE, and educators in South Carolina reviewed the tasks and items for alignment with the ASMGs and for bias and sensitivity concerns. The reviews for content and bias and sensitivity were combined due to the direct impact of the task format, materials, and language on the assessment accessibility for the population. Committees comprising teachers of students with significant cognitive disabilities, representatives of higher education, special education administrators, experts in the instruction of students with limited English proficiency, and content experts from across the state participated in these reviews to consider the following:

- Alignment to the ASMGs and Extended Standards
- Bias for specific groups and types of disabilities
- Accessibility of the tasks to the entire population for whom the test was designed
- Characteristics that might lead to bias or are inappropriate for or insensitive to the nature of the student subgroups (e.g., exclusionary language, stereotypes)
- Format and content of the tasks
- Accessibility of materials
- Clarity of instructions and ease of administration

The review committee meetings were conducted in November 2005, May 2006, and, for the spring 2009 embedded field test, in November 2008. For the 2010 biology field test, the content and bias and sensitivity review meetings were held in June 2009. During the reviews, committee members recommended that some items be revised or eliminated.

Development of Field-Test Tasks and Forms

- On the basis of the feedback from all the steps above, AIR conducted a final review and sign-off for all items and tasks. Following this review, the items and tasks were affirmed ready for field-testing.
- Prior to assembling tasks into test forms, the senior content lead for each content area and the project director reviewed the items and tasks one last time to determine whether the revisions were appropriate and maintained the alignment of the item to the targeted standard.
- For stand-alone field tests, tasks and their items were then placed into field-test forms consistent with the specifications described earlier. For embedded field tests, the tasks and their items were placed into designated locations on the operational test forms.

Item Data Review

- After field-testing, AIR and SCDE staff, including alternate assessment specialists, psychometricians, content specialists, and special educators, met to review the field-test statistics.
- They reviewed the statistics associated with each item and task to determine whether the items were functioning within expectations and whether the tasks were appropriately placed within the instrument. The statistical criteria applied to the field-test item data and to the operational item data are described in Chapter 5.
- The committee also considered teacher comments on specific items from the field test, data from field-test observations, and the results of the alignment studies to make decisions about the inclusion of items in the operational assessment.
- Items that did not meet these criteria were retained for possible future operational use or were revised for recalibration.
- The Item Data Review meetings for the original independent field tests were conducted in August 2006 and June 2007, and an Item Data Review for the spring 2008 embedded social studies field test was held in connection with the planning meeting in September 2008. The Item Data Review meeting of the 2009 embedded field-test items in ELA, mathematics, science, and social studies was conducted in July 2009. The Item Data Review of the 2010 independent biology field test was held in July 2010.

Development of Operational Task/Item Pool

- AIR once again reviewed all data associated with the tasks and items to determine whether the items were functioning as expected and were useful for measuring the achievement of students in South Carolina.
- Items that survived all review and analysis criteria were placed into the operational task/item pool.

Design and Development of the 2006–2010 SC-Alt Field Tests

Following the task development process, the field-test forms were designed and produced. The primary purposes of the independent field-test administrations for English language arts and mathematics (spring 2006), science (fall 2006), and social studies (spring 2007) were to produce data to evaluate SC-Alt tasks and items and to guide the assembly of operational test forms to be used in 2007 and beyond. Student scores based on field-test data were not reported.

An embedded field test (spring 2008) tested the symbolate version of social studies task "George Washington", so that its performance could be compared against the text version used in the spring 2007 field test.

The design, data collection, and analysis of the independent 2006 and 2007 field tests in ELA, mathematics, science, and social studies, of the 2008 embedded social studies field-test tasks, and of the 2009 embedded field-test tasks in ELA, mathematics, science, and social studies were discussed in the spring 2007, 2008, and 2009 operational technical reports.

Spring 2010 Independent Field Test in Biology

During spring 2010, concurrently with the operational SC-Alt administration, 21 new high school biology tasks were field-tested on 15- and 16-year-old students eligible for alternate assessment. The biology field test was administered to 472 students on two forms of 12 tasks each. The forms were linked by three shared tasks, which allowed all biology items to be calibrated on the same scale.

The IRT parameters, classical item statistics, and fit and DIF statistics were subjected to an item data review conducted with AIR and SCDE staffs, July 20, 2010. A standard-setting workshop based on the biology field-test data was conducted September 14–15, 2010.

Use of the Student Placement Questionnaires

The Student Placement Questionnaires (SPQs) are brief structured rating instruments that represent the range of communication levels and cognitive-academic functioning found in the population of alternate assessment examinees. AIR developed the SPQ for the South Carolina Alternate Assessment program.

The student placement process is intended to achieve several important goals:

- It matches student achievement levels with the difficulty of the tasks and items that are administered.
- It allows a maximum number of student item responses at an appropriate level of difficulty.
- It minimizes fatigue by targeting the assessment to the student.
- It supports the psychometric rigor of student scores. A student is administered a better targeted test than one that contains many items the student might find too difficult. Better test targeting contributes to better score reliability. Because fatigue effects from the student's limited attention span are reduced, validity of the overall assessment is enhanced.

Teachers completed the SPQs in each content area to identify the most appropriate starting task for each student. For each subject, the SPQs prompted the teacher with between 12 and 15 "can do" questions (e.g., can this student recognize the sun, moon, Earth?). The questions were grouped by major content standards and sampled across low-, moderate-, and high-complexity levels. Each question rated the student's functioning on a 4-point scale, valued 0 to 3. Answering the 12 to 15 questions of each SPQ, summing the total score, and identifying the most appropriate starting task in a look-up table took test administrators approximately 6 or 7 minutes.

The look-up table identified ranges of SPQ scores that corresponded to one of three starting tasks. Teachers used the SPQs to assign students to starting points on the assessment. Cut points for the science SPQ were based on the rules derived for the mathematics SPQ but were altered for the number of items on the science SPQ. Details regarding the student participation, analysis, and conclusions drawn from use of the SPQ placement procedure appear below.

Administration: Placement and Stopping Rules

After teachers identified the most appropriate starting task for a student, they followed several rules as they administered the starting task and subsequent tasks. If starting at task 1, the teacher would administer at least five tasks; otherwise, at least seven tasks would be administered. For detailed placement and stopping rules for the spring 2010 operational and field-test administrations, see Appendix B.

SPQ Summary

The preceding discussion reviewed some of the implementation procedures for the SPQ. Here we review two of the technical characteristics of the SPQ: the method used to select the SPQ recommended starting task and the usefulness of the SPQ as an indicator of student starting task.

The technical development of the SPQ and determination of the cut points to determine starting tasks are fully described in American Institutes for Research, 2008, *South Carolina Alternate Assessment (SC-Alt): Technical Report for English Language Arts and Mathematics Field Test Administration, Spring 2006*.

Usefulness of the SPQ for Determining the Starting Task. AIR has gathered information regarding the agreement between the SPQ recommended start points and the final observed start points by reviewing item data following each of the 2007, 2008, 2009, and 2010 administrations. The results of the study of 2010 data are reported in detail in Chapter 8.

Use of the SPQ pre-assessment score is only the first step in the procedure used by the test administrator in determining where the student should start the assessment. The instructions for using the SPQ include procedures requiring teachers to adjust the starting point below the SPQ recommended start point when the student is not successful on the first administered task. Alternately, after reviewing the assessment, some teachers may have judged that a student needed to start at a higher level than recommended by the SPQ.

The results of the 2010 study indicate that the agreement between the SPQ recommended start point and the observed start point by content area were 97% for ELA, 95% for mathematics, 96% for science, and 96% for social studies. Since the test administrator is required to make adjustments based on the student's success on the first task, and these adjustments are reflected in the agreement rates, the SPQ appears to be working very effectively for targeting the first task to begin the assessment process.

The results of the Start-Stop Analysis reported in Chapter 8 also support the effectiveness and validity of the SPQ and the SC-Alt tailored assessment design.

Teacher Scoring Accuracy

The design of the SC-Alt includes test administrator (teacher) scoring of student responses. The degree of accuracy with which the test administrator evaluates student performance determines whether the student receives the correct scores and the correct performance level.

A video study was conducted during the 2009 administration to confirm that test administrators were following all scoring procedures accurately. For this study, scoring accuracy refers to the degree to which teachers follow scaffolding and scoring directions correctly and assign correct

scores to student responses. Scoring accuracy by the test administrators was evaluated by having trained raters at AIR review the videotapes of the test administrations and score the student responses without knowledge of the scores assigned by the test administrators. After the raters concluded their scoring of the student responses, the consistency between the test administrators and AIR raters was determined.

The scoring consistency analysis was presented in the 2009 SC-Alt Technical Report, Appendix C. The results indicated that there was a high degree of consistency between the scoring of the test administrators and the AIR raters, suggesting that test administrators in South Carolina understood the scoring procedures and implemented them accurately when scoring student responses.

2010 Operational Test Booklets and Administration and Scoring Procedures

For each grade-band test form in each content area, tasks and items were selected that met the statistical criteria and that covered the breadth of the targeted ASMGs or Extended Standards. The 2010 operational test forms in ELA, mathematics, science, and social studies were revised from the 2009 operational administration. All operational forms had their tasks ordered by increasing difficulty of the items in each task, which was determined by Item Response Theory (IRT) analysis. The goal was to use technically sound assessment instruments to support valid inferences about what students know and can do relative to the ASMGs or Extended Standards in each content area.

The SC-Alt operational administration in spring 2010 included three sets of test materials in English language arts, mathematics, and science: one for the 3–5 grade-band assessment, one for the 6–8 grade-band assessment, and one for the grade 10 assessment. The social studies assessment used two sets of materials for grade-bands 3–5 and 6–8 (grade 10 is not part of the social studies assessment). Teachers (test administrators) received a *Test Administration Manual* (*TAM*) and comprehensive training based on the manual and the test materials.

The 2010 test booklets contained 12 operational tasks in each content area. Operational tasks were arranged in test forms in the order of the empirical difficulty of the first item in each task. Each test form (elementary, middle, and high school) included linking tasks to support psychometric linking of the grade-band score scales. Each task consisted of four to eight separate items. Teachers were instructed to administer a minimum of five or seven tasks to each student, depending on the SPQ designated starting point, and to continue administration of subsequent tasks until the student was no longer successful.

Teachers also received other materials with each test booklet:

- A physical manipulatives kit
- Printed manipulatives
- An answer folder for each participating student
- A Student Placement Questionnaire and directions for determining the starting task for each student

Exhibit 2.1 summarizes the operational grade-band assessments and the numbers of tasks in each grade assessment for 2010.

Exhibit 2.1: Numbers of Operational and Field Test Tasks in Each Grade-Band Assessment, 2010

Grade-Band	Total in Each Grade-Band (Field-Test Tasks in Parentheses)				
	ELA	Math	Science	Social Studies	Biology
3–5	12	12	12	12	
6–8	12	12	12	12	
10	12	12	12		(21, FT)

The approximate test length for each grade-band assessment for the 2010 administration was 60 items (12 tasks × an average 5 items per task) and 120 score points (60 items × an average 2 points per item).

Linking Tasks in Each Grade-Band Assessment

All tasks in each SC-Alt grade-band assessment align with the original ASMGs in that grade-band. For example, the first two items in Task 9, which is part of the grade-band 3–5 ELA assessment (see Exhibit 2.2), align with Measurement Guideline 27, "Identify the problem and/or the solution in a story or drama," and Measurement Guideline 16, "Identify the impact of a given cause or effect on a given character." These Measurement Guidelines are linked to State Academic Standards and Indicators 3-R2.2 and 5-R1.11, respectively, from grade-band 3–5.

Because adjacent grade-band score scales are linked psychometrically, some tasks in each grade-band assessment align with the original ASMGs in both adjacent grade-bands. For example, ELA Task 10, which provides data for psychometric linking of the grade-bands 3–5 and 6–8 score scales, aligns with ASMGs at both grade-bands. Item 1 in Task 10 aligns with ASMG 30, "Identify the purpose of a text"; that ASMG is linked to State Academic Indicator 3-R2.10 at grade-band 3–5 and State Academic Indicator 6-R2.9 at grade-band 6–8 (see Exhibit 2.2).

All items in linking tasks are designed to be appropriate for students in both adjacent grade-bands. The alignment studies (discussed in Chapter 8) confirm that all tasks in each grade-band, including linking tasks, align with ASMGs for each separate grade-band and with the corresponding grade-band academic content standards. In addition, the corresponding grade-level State Academic Standards and Indicators to which the ASMGs are linked do differ across the adjacent grade-bands. (See Appendix A.)

Exhibit 2.2: Two Tasks from the Grade-Band 3–5 ELA Assessment that Illustrate the Alignment of Items of Grade-Band ASMGs and State Academic Standards

Item	SC-Alt ASMG	Corresponding State Academic Standards and Indicators from Grades 3–5	Corresponding State Academic Standards and Indicators from Grades 6–8
Task 9			
Item 1	Recognize conflict in stories: • 27. Identify the problem and/or the solution in a story or drama.	3-R2.2: Demonstrate the ability to identify problem and solution in a work of fiction or drama.	--
Item 2	Determine cause and effect in texts read aloud or independently: • 16. Identify the impact of a given cause or effect on a given character.	5-R1.11: Demonstrate the ability to analyze cause and effect.	
Item 3	Analyze plots, characters, and settings in literature: • 25. Identify and describe characters, settings, and events in a story.	3-R2.1: Demonstrate the ability to analyze characters, setting, and plot in a literary work.	
Item 4			
Item 5	Make predictions about text: • 9. Use pictures and words to make predictions about texts read aloud or independently.	3-R1.7: Demonstrate the ability to make predictions about stories.	
Task 10			
Item 1	Identify the author's purpose: • 30. Identify the purpose of a text.	3-R2.10: Continue identifying the author's purpose in a variety of texts.	6-R2.9: Demonstrate the ability to identify the author's purpose in texts in a variety of genres.
Item 2	Recall details: • 7. Recall details in tests read aloud or independently. Follow written directions: • 17. Follow written one-step or multistep directions (presented in words/pictures/ symbols/ objects).	3-R1.5: Demonstrate the ability to recall details in texts. 3-R1.13: Demonstrate the ability to follow a logical sequence of written directions to complete a task.	6-R1.5: Demonstrate the ability to identify the details that support the thesis of a particular text. 6-R1.11: Demonstrate the ability to follow multistep directions such as those for preparing applications and completing forms.
Item 3			
Item 4			
Item 5			

Item	SC-Alt ASMG		Corresponding State Academic Standards and Indicators from Grades 3–5		Corresponding State Academic Standards and Indicators from Grades 6–8
Item 6	Make predictions about text: • 9. Use pictures and words to make predictions about texts read aloud or independently.		3-R1.7: Demonstrate the ability to make predictions about stories.		6-R1.6: Demonstrate the ability to make predictions about stories.

Chapter 3: Spring 2010 Operational Test Administration

This section describes the spring 2010 operational test administration in the following areas:

- Student participation for the spring 2010 administration
- Demographics of participating students
- Test administration window, materials, and timelines
- Test administrator requirements
- Test administrator training
- Pre-assessment using the Student Placement Questionnaire
- Fidelity of administration and accuracy of scoring
- Test security provisions

Student Participation for the Spring 2010 Administration

Students participating in the spring 2010 operational administration were those students whose IEP team had determined that they met the following SC-Alt participation criteria for alternate assessment and who were ages 8–13 or 15 on September 1, 2009. These are the ages of typical students who are in grades 3–8 and 10.

- The student demonstrates a significant cognitive disability and adaptive skills, which result in performance that is substantially below grade-level achievement expectations even with the use of accommodations and modifications.
- The student accesses the state-approved curriculum standards at less-complex levels and with extensively modified instruction.
- The student has current adaptive skills requiring extensive direct instruction and practice in multiple settings to accomplish the application and transfer of skills necessary for application in school, work, home, and community environments.
- The student is unable to apply or use academic skills across natural settings when instructed solely or primarily through classroom instruction.
- The student's inability to achieve the state grade-level achievement expectations is not the result of excessive or extended absences or social, cultural, or economic differences.

Exhibit 3.1 indicates the age ranges of students who participated in the SC-Alt in spring 2010.

Exhibit 3.2 indicates the alternate assessment eligibility categories that were placed in each eligible student's state precoding file (precoding files enabled SCDE and AIR to ensure that the appropriate SC-Alt materials were delivered to teachers in time for the spring 2010 administration).

Exhibit 3.1: Age Reference Sheet for 2009–10 Alternate Assessment, Spring 2010 Operational Administration

Age as of 9/1/09	Corresponding Birth Date Range		Test Required 2009–10	Precode AA Eligibility Code
	Beginning DOB	Ending DOB		
5	09/02/03	09/01/04	none	5
6	09/02/02	09/01/03	none	5
7	09/02/01	09/01/02	none	5
8	09/02/00	09/01/01	SC-Alt Elem	2
9	09/02/99	09/01/00	SC-Alt Elem	2
10	09/02/98	09/01/99	SC-Alt Elem	2
11	09/02/97	09/01/98	SC-Alt Middle	3
12	09/02/96	09/01/97	SC-Alt Middle	3
13	09/02/95	09/01/96	SC-Alt Middle	3
14	09/02/94	09/01/95	none	5
15	09/02/93	09/01/94	SC-Alt HS	4
16	09/02/92	09/01/93	none	5
17	9/02/91	09/01/92	none	5
18	9/02/90	09/01/91	none	5
19	9/02/89	09/01/90	none	5
20	9/02/88	09/01/89	none	5
21	9/02/87	09/01/88	none	5

Exhibit 3.2: Precode Project Coding (Alternate Assessment Eligibility Field)

Code	SASI Dropdown List Description	Full Description
0	Criteria not met	The student does not meet criteria for alternate assessment.
2	**SC-Alt Elem School**	The student requires alternate assessment and meets the age eligibility requirement for assessment with the **SC-Alt Elem School form** this current school year (8–10 years old on September 1, 2009).
3	**SC-Alt Middle School**	The student requires alternate assessment and meets the age eligibility requirement for assessment with the **SC-Alt Middle School form** this current school year (11–13 years old on September 1, 2009).
4	**SC-Alt High School**	The student requires alternate assessment and meets the age eligibility requirement for assessment with the **SC-Alt High School form** this current school year (15 years old on September 1, 2009).
5	**AltAssess NotAgeElig**	The student requires alternate assessment, **but does not meet the age eligibility requirements to be assessed with SC-Alt this current school year** (i.e., the student was younger than eight years, age 14, or older than 15 years on September 1, 2009).

Demographics of Participating Students

This section describes the demographics of participating students by test form (elementary, middle, or high school). Exhibit 3.3 presents the student demographics for participating students in each grade-band.

For the purpose of this report, the inclusion of students was based on the same criteria applied in the reporting of student scores. A student was included if the following criteria were met: (1) a signed security affidavit was received for the student, (2) the student was not noted to be excluded from reporting for some other reason (e.g., inappropriate administration procedures), and (3) the number of coded responses met the attemptedness requirement for student scoring (i.e., five valid responses) in at least one content area. The population of students reported, therefore, includes 1,393 elementary school test forms, 1,185 middle school test forms, and 357 high school test forms.

According to the attemptedness requirements, a student's responses to a test form could be assigned to one of four completion status categories: completion ("student satisfied attemptedness rule"), invalid due to too few scored responses ("student did not satisfy attemptedness rule"), invalid due to test administration errors ("test administrator did not follow instructions for starting tasks") or not tested ("student did not answer any content area items"). For all content areas, the majority of students reported completed the administered test form; 99% or more of the eligible students completed ELA and mathematics, 68% to 69% completed science and social studies in the elementary and middle school grade-bands[1], and 99% completed the high school science assessment. Of the remaining student records, less than 1% of reported test forms were categorized as not tested or not meeting the attemptedness criteria.

Given that the number of students to be assessed on the high school test form was approximately one-third the number of students assessed on either the elementary or the middle school forms, the proportion of demographic characteristics of the student population was relatively consistent across grade-bands. In terms of ethnicity, African American students made up 48% to 53% of the assessed students across grade-bands; white students accounted for 41% to 43% of the students across grade-bands, and Hispanic students accounted for 3% to 5% of students across forms. Other ethnicities each accounted for less than 2% of the assessed population. Gender was also relatively consistent across grade-bands, showing a slight proportional decrease of males from elementary and middle school to high school, with approximately a two-to-one ratio of male students (66% to 67%) to females (33% to 34%).

The classification of students in terms of English language proficiency was also consistent across grade-bands. The majority of students (95% to 97%) were classified as "English Speaker II," meaning that they had never been coded as an ESL student. The remaining language proficiency classifications each accounted for less than 2% of students by grade-band with the exception of "Pre-functional" (2% to 4%), indicating that the student scored pre-functional on the English language proficiency assessment and is receiving English as a second language (ESL) services. The percentage of pre-functional ESL students decreased across grade-bands.

The grade reported for a student in the school's database is the grade reported for funding purposes (EFA grade) and is often determined by the location of the student's educational

[1] Not all students were required to complete the science and social studies subject areas.

program instead of by the student's age or years in school. Therefore, approximately 9% of students administered the elementary form (for students ages 8–10, the typical ages of students in grades 3–5) had reported EFA grades lower than grade 3 or higher than grade 5, with most of these students classified in the contiguous grades of 2 and 6. Of students administered the middle school form (for students ages 11–13, the typical ages for grades 6–8), 22% of the students were reported at grades below grade 6 or above grade 8. The vast majority of these students were classified as grade 5 students (18% of all middle school form students), which indicates that these students were being served in educational programs housed in elementary schools. Of the students administered the high school form (for students age 15), 74% were reported as grade 9 or grade 10 (35% and 39%, respectively). Sixteen percent (16%) of the high school form students were reported as grade 8 students, indicating that these students were being served in educational programs housed in middle schools. The purpose of assigning SC-Alt grade-band forms by age is to ensure that students are instructed and assessed on the appropriate grade-band curricula regardless of where their educational programs are housed.

The percentage of students receiving free lunch at schools decreases slightly across forms (64% to 59%), and the percentage of students receiving reduced-price meals is approximately the same across forms (8% to 9%). One student was indicated as being a migrant student; another student was indicated as being home-schooled. Twenty-three elementary school students (less than 2%) were indicated as being medically homebound, as were 11 middle school students (1%) and eight high school students (2%).

Fourteen different disability codes were reported for students assessed with the SC-Alt. The coding system allowed students to be coded with more than one disability code. Students with the primary disabilities of severe mental disability, moderate mental disability, mild mental disability, and autism made up 86% to 90% of the students assessed with the SC-Alt (86%, 89%, and 90% for the elementary, middle school, and high school forms respectively). Of these, the percentage of students coded as having moderate mental disability increases across test forms (24% to 45%), while mild mental disability and autism decrease across test forms and the percentage of severely mentally disabled stays about the same (approximately 11%). Although a few students would have a primary disability of speech or language impairment, the vast majority of students received this code because they were receiving speech/language therapy as a supplementary service.

Exhibit 3.3: Summary of Demographic Information

	Elementary		Middle		High	
	N	%	N	%	N	%
Student's Ethnicity						
African American	671	48.17	619	52.24	190	53.22
African American/American Indian	8	0.57	4	0.34	.	0.00
American Indian	5	0.36	3	0.25	.	0.00
Asian	11	0.79	14	1.18	4	1.12
Hawaiian/Pacific Islander	1	0.07	.	0.00	.	0.00
Hispanic	66	4.74	43	3.63	10	2.80
White	593	42.57	486	41.01	150	42.02
White/African American	22	1.58	11	0.93	3	0.84

	Elementary		Middle		High	
	N	**%**	**N**	**%**	**N**	**%**
White/American Indian	1	0.07	1	0.08	.	0.00
White/Asian	6	0.43	1	0.08	.	0.00
Other	9	0.65	3	0.25	.	0.00
Unknown	.	0.00	.	0.00	.	0.00
Student's Gender						
Female	455	32.66	391	33.00	123	34.45
Male	938	67.34	794	67.00	234	65.55
Unknown	.	0.00	.	0.00	.	0.00
ESL (Language)						
Advanced	.	0.00	.	0.00	.	0.00
Advanced waiver	.	0.00	.	0.00	.	0.00
Beginner	1	0.07	1	0.08	.	0.00
Beginner waiver	.	0.00	.	0.00	.	0.00
English speaker I	.	0.00	.	0.00	.	0.00
English speaker II	1317	94.54	1136	95.86	348	97.48
Full English proficient	1	0.07	1	0.08	1	0.28
Intermediate	.	0.00	2	0.17	.	0.00
Intermediate waiver	.	0.00	.	0.00	.	0.00
Pre-functional	50	3.59	31	2.62	7	1.96
Pre-functional waiver	2	0.14	1	0.08	.	0.00
Title III first year exited	.	0.00	.	0.00	.	0.00
Title III second+ year exited	.	0.00	.	0.00	.	0.00
Unknown	22	1.58	13	1.10	1	0.28
Eligible for Free or Reduced-Price Lunch						
Free meals	894	64.18	724	61.10	212	59.38
Full-pay meals	390	28.00	354	29.87	113	31.65
Reduced meals	109	7.82	107	9.03	31	8.68
Unknown	.	0.00	.	0.00	1	0.28
EFA Grade (Reported Grade for Funding)						
1	6	0.43	.	0.00	.	0.00
2	90	6.46	.	0.00	.	0.00
3	510	36.61	7	0.59	.	0.00
4	443	31.80	21	1.77	.	0.00
5	320	22.97	218	18.40	7	1.96
6	14	1.01	378	31.90	2	0.56
7	5	0.36	344	29.03	10	2.80
8	4	0.29	200	16.88	56	15.69
9	1	0.07	14	1.18	126	35.29
10	.	0.00	2	0.17	139	38.94

	Elementary		Middle		High	
	N	%	N	%	N	%
11	.	0.00	1	0.08	12	3.36
12	.	0.00	.	0.00	5	1.40
Completion Status: Attempted						
ELA	1391	99.86	1184	99.92	357	100.00
Math	1386	99.50	1183	99.83	356	99.72
Science	949	68.13	810	68.35	355	99.44
Social Studies	948	68.05	821	69.28	.	0.00
Completion Status: Not Tested						
ELA	2	0.14	1	0.08	.	0.00
Math	6	0.43	1	0.08	1	0.28
Science	444	31.87	375	31.65	2	0.56
Social Studies	444	31.87	362	30.55	357	100.00
Completion Status: Test not Valid—Student Received Fewer Than Five Scored Responses						
ELA	.	0.00	.	0.00	.	0.00
Math	.	0.00	.	0.00	.	0.00
Science	.	0.00	.	0.00	.	0.00
Social Studies	1	0.07	1	0.08	.	0.00
Completion Status: Test not Valid—Test Administrator Did not Follow Instructions for Starting Tasks						
ELA	.	0.00	.	0.00	.	0.00
Math	1	0.07	1	0.08	.	0.00
Science	.	0.00	.	0.00	.	0.00
Social Studies	.	0.00	1	0.08	.	0.00
Special School Status Fields						
Migrant status	1	0.07	.	0.00	.	0.00
Home schooled	.	0.00	1	0.08	.	0.00
Medical homebound	23	1.65	11	0.93	8	2.24
IEP Disability Codes (Multiple Codes per Student)						
Severely mentally disabled	147	10.55	135	11.39	38	10.64
Moderately mentally disabled	339	24.34	409	34.51	161	45.10
Mildly mentally disabled	387	27.78	274	23.12	64	17.93
Autism	321	23.04	235	19.83	57	15.97
Deaf/blindness	11	0.79	9	0.76	5	1.40
Emotional disability	9	0.65	4	0.34	2	0.56
Hearing impaired	31	2.23	15	1.27	5	1.40
Learning disability	24	1.72	17	1.43	2	0.56
Multiple-disability	110	7.90	77	6.50	17	4.76
Other health impaired	84	6.03	61	5.15	19	5.32
Orthopedically impaired	85	6.10	51	4.30	13	3.64
Speech or language impaired	1038	74.52	607	51.22	124	34.73

	Elementary		Middle		High	
	N	**%**	**N**	**%**	**N**	**%**
Traumatic brain injury	9	0.65	6	0.51	1	0.28
Visually impaired	50	3.59	47	3.97	14	3.92
Total	**1393**	**100.00**	**1185**	**100.00**	**357**	**100.00**

Test Administration Window, Materials, and Timelines

The spring 2010 administration of the SC-Alt included the following important dates:

- SC-Alt test administration training for teachers new to the SC-Alt operational administration (did not administer in 2007, 2008, or 2009), five regional SCDE workshops, January 11–15, 2010
- District-level SC-Alt test administration training for all test administrators, February 1–26, 2010
- Test materials arrived in district, February 24, 2010
- Assessment window, March 8–April 30, 2010
- Teachers returned materials to DTC-Alt, May 3, 2010
- Materials received by contractor, May 7, 2010

Teachers had approximately eight weeks to review the materials and complete the test administration. Teachers received both printed and physical manipulatives to use during test administration. They were also responsible for collecting a few common classroom items that were familiar to the student to use with several tasks.

Test Administrator Requirements

Test administrators are required to receive training on all phases of the administration of the SC-Alt and must be one of the following:

- A certified employee of the district
- An employee of the district who is a critical needs teacher and has a letter of eligibility, an interim certificate, or a critical needs certificate
- A substitute teacher who is certified and employed by the district on an as-needed basis
- Someone who was a certified teacher but has allowed the teaching certificate to expire owing to retirement, change of career, or some other reason and has been approved by the district test coordinator or the DTC-Alt as a qualified test administrator
- Someone who is not certified but has been employed by the school district in an instructional capacity and has been approved by the DTC-Alt as a qualified test administrator

If a test is administered in a location other than the school, the test administrator must still meet the criteria specified above.

Test Administrator Training

Test administration training is required for all test administrators. The SC-Alt is individually administered with a standard script and scored by the test administrator as the assessment is conducted. Fidelity of administration and scoring is essential to the validity of the assessment results.

Teachers who administered the SC-Alt during the spring of 2010 but who did not administer the SC-Alt in spring 2007, 2008, or 2009 were required to attend a South Carolina Department of Education (SCDE) training session. In addition, all teachers who administered the SC-Alt in spring 2010, including those who attended the SCDE workshops, were required to attend a district-level SC-Alt administration training session conducted by the district test coordinator for alternate assessment (DTC-Alt). At the completion of the training sessions, each test administrator was required to sign and submit to SCDE an acknowledgment of receiving training and readiness to conduct the assessment.

The training included the following elements:

- Review of the eligibility criteria for students participating in the alternate assessment
- Overview of the ASMGs and Extended Standards, emphasizing the link to the general education standards
- Explanation of how the assessment was developed, including the role of the review committees
- Review of test administrator requirements, test security, and test materials
- Training and practice in pre-assessment using the SPQ
- Description of the assessment format and procedures:
  - Setup
  - Script
  - Scoring
  - Adaptive instructions
- Instruction for making SC-Alt tasks accessible
- Overview of assistive technology and the alternate assessment
- Administration and scoring instruction and practice using released test items provided on video clips of South Carolina teachers actually administering a task to students representing a variety of disabilities and ethnicities
- Scoring qualifying round
- Review of procedures for receiving and shipping materials back to the DTC-Alt

Pre-Assessment Using the Student Placement Questionnaire

As noted earlier in this Technical Report, the administration of the SC-Alt uses the SPQ as a pre-assessment instrument to determine the most appropriate starting point in the assessment. Recall that the SPQ requires the teacher to evaluate the student on 12 to 15 "can do" statements addressing the student's skills and knowledge in each content area on the basis of the teacher's prior instructional knowledge of the student. A total score computed from the teacher's SPQ responses indicates the initial starting task for the assessment. Once the assessment has begun, the test administrator is required to adjust the starting point for the student if the student is not successful on the first task. Rules have been established for adjusting the starting tasks and for

determining when the assessment should be concluded. The starting and stopping rules used with the SPQs for the 2010 administration are presented in Appendix B.

Fidelity of Administration and Accuracy of Scoring

During the assessment administration, a monitor must be present to observe all assessment sessions and verify the use of proper assessment procedures and the authenticity of student responses. Monitors must be trained and sign a Test Administrator Security Affidavit to verify that the appropriate procedures were used. The Test Administrator Security Affidavit contains a coded label to link the document to the student answer folder and includes the principal's verification of the use of appropriate assessment and scoring procedures. If this document is missing, the administration is considered an invalid administration.

Test Security Provisions

This section describes the test security procedures associated with the SC-Alt. SCDE has the following test security measures in place:

- Each local school board must develop and adopt a district test security policy. The policy must provide for the security of the materials during testing and the storage of all secure tests and test materials before, during, and after testing. Before and after testing, all materials must be stored at a location(s) in the district under lock and key.
- Each District Superintendent must designate annually one individual in each district for each mandated assessment who will be the sole individual in the district authorized to procure test instruments that are used in testing programs administered by or through the State Board of Education. The designated individual for alternate assessment is the district test coordinator for alternate assessment (DTC-Alt). The DTC-Alt is responsible for receiving and distributing all SC-Alt materials and ensuring that all SC-Alt administration procedures and requirements are met.
- All school and district personnel who may have access to SC-Alt test materials or to the location in which the materials are securely stored must sign the Agreement to Maintain Test Security and Confidentiality before they are given access to the materials.
- Test administrators must be trained annually to administer the SC-Alt and must meet all test administrator requirements.
- An assessment monitor must observe all assessment sessions and verify the use of proper assessment procedures and the authenticity of student responses for each completed assessment.
- Test administrators must complete an SC-Alt Test Administrator Security Affidavit for each student they assess and submit the form with the student's assessment materials.

Chapter 4: Setting Performance Standards

For the South Carolina Alternate Assessment, two standard-setting workshops were conducted: One, in June 2007, convened a diverse panel of 105 educators, parents and educational administrators to recommend status performance standards based on the spring 2007 operational test administration data for ELA, mathematics, and science and field-test data for social studies; in the second workshop, in September 2010, 19 panelists recommended standards for high school biology based on spring 2010 field-test data. This chapter summarizes the descriptions of achievement levels, the procedures used for setting standards for each content area, and the recommended standards themselves, including student impact information. Complete details of the two standard-setting workshops can be found in separate reports (American Institutes for Research, 2007; American Institutes for Research and South Carolina Department of Education, 2010).

Using the Item Descriptor (ID) Matching method (see Cizek & Bunch, 2007; Ferrara, Perie, & Johnson, 2008), the panelists reviewed test items and the corresponding Descriptions of Achievement Levels (DALs) and then recommended performance standards for Level 2, Level 3, and Level 4 achievement levels. These standards were then translated into cut points on the student proficiency scale by AIR psychometricians. This section describes the process and outcomes of the standard-setting workshop.

Descriptions of Achievement Levels

DALs are key elements in standard-setting processes. DALs define the content area knowledge, skills, and processes that examinees at a performance level are expected to possess. The descriptions of Level 1, Level 2, Level 3, and Level 4 performance that SCDE developed make up the public statement about what and how much South Carolina educators want students to know and be able to do for each grade level and content area. Level 3 and higher represents "proficient performance" for NCLB reporting.

The development of the DALs for ELA, mathematics, science and social studies followed a multistep process involving AIR staff and SCDE staff working with committees of teachers, parents, and special education administrators. The process was begun by examining the DALs used with the other South Carolina assessment programs (PACT, HSAP, PACT-Alt, and HSAP-Alt) and the performance-level descriptors for alternate assessments used by other states. During spring 2007, these DALs were developed and refined over multiple meetings between AIR, SCDE and stakeholder committees to determine what proficiency meant for students participating in each grade-band of the SC-Alt. Some additional refinement occurred during the standard-setting workshop in June 2007; the final version of these DALs was presented to the State Board of Education on September 12, 2007 and posted on the SCDE website.

The DALs for high school biology were written by AIR and reviewed by SCDE prior to the standard-setting meeting. A subcommittee of standard-setting panelists reviewed the biology DALs on the first day of the September 2010 standard-setting workshop.

In the SC-Alt standard-setting workshops, panelists used the DALs presented in Appendix D when they placed their cut scores.

The ID Matching Standard-Setting Process

The ID Matching standard-setting process, described in the standard-setting plans submitted to SCDE and reviewed by the South Carolina Technical Advisory Committee, was used at both standard-setting workshops in Columbia, SC (in June 2007 and September 2010). When standards were to be set in multiple subjects, the panels were divided into subject-specific groups. For subjects that were assessed in multiple grade-bands, anchor standards were first established in the lowest and highest grade-bands (e.g., grades 3–5 and grade 10). AIR staff provided training and led the panelists through two rounds of ID Matching to set the Level 3 standard first, followed by the Level 2 and 4 standards.

Before the participants made each of their recommendations using the ID Matching procedure, they were given a readiness form to ensure that they fully understood the task and were prepared to place the performance standard. The participants indicated unanimously that they understood the task and were prepared to make performance standard recommendations.

Goals of the Standard Setting

The goals of the meeting, as stated to the panelists, were as follows:

- Recommend performance standards on the ELA, mathematics, science, social studies, and/or biology assessments that correspond to the DALs for Level 2, Level 3, and Level 4 performance levels
- Consider the agreement and impact data to guide judgments about item difficulty and placement of the performance standards
- Recommend to SCDE the appropriate placement of cut points on the student proficiency scales for each grade-band assessment

Panel Composition

Across the two workshops, 124 panelists participated in recommending performance standards across five content areas: ELA, mathematics, science, social studies, and biology. The overall composition of the panel followed the SCDE-provided specifications and was broadly designed to ensure that the panel was widely diverse and represented a cross-section of South Carolina's educators and non-educators.

Standard-Setting Workshop Activities

Workshop participants recommended performance standards for the assessments during two rounds of deliberation for each DAL in each content area and in each grade-band as follows.

- Set standards in anchor grade-bands (3–5 and 10)
    - Participants complete Rounds 1 and 2 for each performance-level standard.
    - Table leaders articulate standards across grades and content areas (align them on the basis of content considerations).
    - For the biology standard setting, table leaders moderate the standards with respect to existing performance standards in science.

- Set standards in intermediate grade-band (6–8), if needed.
  - Participants complete Rounds 1 and 2 for each performance-level standard.
  - Table leaders *articulate* standards across grades and content areas (align them on the basis of content considerations and consistency with anchor grade standards).

Throughout the workshops, the panelists had many opportunities to reflect on the pattern of performance standards they were recommending. Their general conclusion was that they were satisfied that the standards made sense from a content and experiential point of view. They felt that the patterns reflected the requirements of the content standards and the realities of student performance.

With few exceptions, panelists recommended standards that followed an orderly progression of increasing achievement across levels and grade-bands. Specifically, with the exception of mathematics at the grade-band 6–8 and grade 10, all recommended achievement-level standards increased in difficulty in subsequent grade-bands. Exhibits 4.1–4.4 show the scale score associated with the cut score recommended by each panel. These results were achieved through the process of setting cut scores at anchor grades, making sure they resulted in consistent expectations across grade-bands, and providing articulated standards as a starting point for intermediate grade-bands.

Cut Score Review and the Setting of Final Cut Scores

The results of the standard-setting workshops were presented to the Technical Advisory Committee (TAC) of the Office of Assessment, SCDE, on July 27, 2007. The TAC discussed the results of the standard-setting workshops, reviewed the articulation of the cut scores by grade level, and recommended strategies to the Office of Assessment staff for improving the articulation of the final scores while respecting and maintaining the basic cut score decisions made by the workshop panelists.

Responding to the TAC's suggestions, a committee of Office of Assessment staff examined the scale score articulation and the percentage of students in performance levels by grade and recommended minor adjustments to the original cuts made by the workshop panelists. The adjustments made to each cut score and the resulting final cut scores are presented in Exhibits 4.1–4.4.

In Exhibits 4.1–4.4, the combined standard error of the panelist-recommended cut score (labeled "SE 2007") expresses the joint uncertainty of the IRT-based estimate of the conditional standard error of measurement at the cut score, together with the sampling error of the median agreement per cut score among panelists. The standard error of the median cut score agreement among the panelists, as suggested by Huynh (2003), is listed in column 6. However, two additional details about the standard errors of the median are important to note: First, the standard errors were based on the actual recommended cut scores, and any post hoc adjustment to the cut scores was treated as a constant adjustment. In other words, the adjusted cut score still had the same standard error. Second, the standard errors were initially calculated as standard errors of the page numbers in the ordered-item booklet and then transformed to the scale score metric.

The estimate of the conditional standard error of measurement estimate depends on the set of items used at the time and on the sample of operational item response patterns observed in a

given administration. The entries of the right-most column of Exhibits 4.1–4.4 (labeled "CSEM 2010") display the empirically estimated conditional standard error of measurement at the final, adjusted cut score. This latter estimate is taken from the spring 2010 operational data, computed as the root mean square standard error of the scale score estimates within ±2 scale units of the cut point. The CSEM 2010 values indicate the precision of the current test instrument at the final cut points determined in 2007.

Exhibit 4.1: Panel Recommended and Adjusted Final Cut Scores—ELA

	Panel Recommended Cut Scores		Adjustment to Final Cut Scores			
Performance Level	**Scale Score**	**2007 Combined Standard Error (SE 2007)**	**Level of Adjustment (± SE)**	**Final Cut Scale Score**	**Standard Error of Cut Scale Score**	**2010 Conditional Standard Error of Measurement (CSEM 2010)**
Grades 3–5						
Level 2	403	13.75	None	403	2.96	14.17
Level 3	466	9.54	None	466	1.59	9.64
Level 4	491	12.26	None	491	1.73	10.24
Grades 6–8						
Level 2	417	9.64	None	417	3.81	13.19
Level 3	473	7.99	0.5	477	1.09	9.64
Level 4	501	9.18	None	501	1.45	10.89
Grade 10						
Level 2	429	10.56	None	429	3.38	13.80
Level 3	478	9.11	1	487	0.66	9.57
Level 4	503	9.68	1	514	1.77	10.76

Exhibit 4.2: Panel Recommended and Adjusted Final Cut Scores—Mathematics

Performance Level	Panel Recommended Cut Scores		Adjustment to Final Cut Scores			
	Scale Score	2007 Combined Standard Error (SE 2007)	Level of Adjustment (± SE)	Final Cut Scale Score	Standard Error of Cut Scale Score	2010 Conditional Standard Error of Measurement (CSEM 2010)
Grades 3–5						
Level 2	423	10.22	-1	413	0.66	13.98
Level 3	476	9.59	None	476	0.21	11.05
Level 4	526	14.48	None	526	4.63	13.47
Grades 6–8						
Level 2	425	10.18	None	425	0.50	15.26
Level 3	476	9	1.5	489	0.16	9.87
Level 4	529	10.46	0.5	534	0.74	11.89
Grade 10						
Level 2	434	11.93	None	434	2.19	14.35
Level 3	476	14.76	1.5	498	1.97	9.76
Level 4	528	13.19	1	541	3.82	12.00

Exhibit 4.3: Panel Recommended and Adjusted Final Cut Scores—Science

Performance Level	Panel Recommended Cut Scores		Adjustment to Final Cut Scores			
	Scale Score	2007 Combined Standard Error (SE 2007)	Level of Adjustment (± SE)	Final Cut Scale Score	Standard Error of Cut Scale Score	2010 Conditional Standard Error of Measurement (CSEM 2010)
Grades 3–5						
Level 2	430	10.83	None	430	1.51	15.82
Level 3	474	10.36	-0.5	469	3.25	13.09
Level 4	496	10.38	None	496	0.81	12.23
Grades 6–8						
Level 2	447	9.66	None	447	0.06	13.72
Level 3	484	9.61	0.5	489	0.50	11.62
Level 4	514	11.33	None	514	0.95	12.70
Grades 10						
Level 2	463	11.72	None	463	4.71	13.28
Level 3	492	14.44	1	506	8.13	14.02
Level 4	535	14.78	None	535	1.45	16.96

Exhibit 4.4: Panel Recommended and Adjusted Final Cut Scores—Social Studies

Performance Level	Panel Recommended Cut Scores		Adjustment to Final Cut Scores			
	Scale Score	2007 Combined Standard Error (SE 2007)	Level of Adjustment (± SE)	Final Cut Scale Score	Standard Error of Cut Scale Score	2010 Conditional Standard Error of Measurement (CSEM 2010)
Grades 3–5						
Level 2	423	16.64	None	423	2.98	15.80
Level 3	485	14.39	0.5	492	11.93	13.84
Level 4	549	14	None	549	2.04	15.49
Grades 6–8						
Level 2	439	14.04	None	439	5.96	17.18
Level 3	490	12.58	1.5	503	1.28	13.47
Level 4	560	26.91	None	560	10.57	17.72

The preliminary cut score ranges and associated impact data for the high school biology assessment are given in Exhibit 4.5. These are the cut point ranges recommended by the September 2010 standard-setting workshop, based on independent field-test data. The plan is to re-evaluate the figures of Exhibit 4.5 against operational test data to be collected at spring 2011 administration. At that time, the South Carolina Department of Education will also select a unique cut score for each performance level.

Exhibit 4.5: Panel Recommended Cut Score Ranges—High School Biology

Cut Score Range	Panel Recommended Cut Score Ranges	
	Scale Score	Percent Students at or above Cut Score
Level 2		
Low	404	91.30
Median	407	90.70
High	408	90.49
Level 3		
Low	462	73.00
Median	468	69.90
High	484	60.24
Level 4		
Low	500	49.90
Median	513	42.23
High	519	38.69

For ELA, mathematics, science, and social studies, the final cut scores, the percentage of students performing at each performance level, and the cumulative percentage of students at or above each level are presented in Exhibits 4.6–4.9. These cut scores were approved by the State Superintendent of Education and were presented to the South Carolina State Board of Education on September 12, 2007.

Exhibit 4.6: Percentage of Students at Each Performance Level—ELA

Performance Level	Scale Score Cut Score	Percentage in Level	Cumulative Percentage (at and above) for Each Performance Standard
Grades 3–5			
Level 1	—	12.6 %	100.0 %
Level 2	403	25.4 %	87.4 %
Level 3	466	21.9 %	62.0 %
Level 4	491	40.1 %	40.1%
Grades 6–8			
Level 1	—	12.9 %	100.0 %
Level 2	417	23.3 %	87.2 %
Level 3	477	14.9 %	63.9 %
Level 4	501	49.0 %	49.0 %
Grade 10			
Level 1	—	13.4 %	100.0 %
Level 2	429	23.6 %	86.6 %
Level 3	487	12.5 %	63.1%
Level 4	514	50.6 %	50.6 %

Exhibit 4.7: Percentage of Students at Each Performance Level—Mathematics

Performance Level	Scale Score Cut Score	Percentage in Level	Cumulative Percentage (at and above) for Each Performance Standard
Grades 3–5			
Level 1	—	14.3 %	100.0 %
Level 2	413	30.8 %	85.7 %
Level 3	476	29.3 %	54.9 %
Level 4	526	25.7 %	25.7 %
Grades 6–8			
Level 1	—	15.9 %	100.0 %
Level 2	425	28.5 %	84.1%
Level 3	489	25.9 %	55.6 %
Level 4	534	29.8 %	29.8 %
Grade 10			
Level 1	—	16.1 %	100.0 %

Performance Level	Scale Score Cut Score	Percentage in Level	Cumulative Percentage (at and above) for Each Performance Standard
Level 2	434	30.1 %	84.0 %
Level 3	498	28.9 %	53.9 %
Level 4	541	24.9 %	24.9 %

Exhibit 4.8: Percentage of Students at Each Performance Level—Science

Performance Level	Scale Score Cut Score	Percentage in Level	Cumulative Percentage (at and above) for Each Performance Standard
Grades 3–5			
Level 1	—	19.8 %	100.0 %
Level 2	430	18.2 %	80.2 %
Level 3	469	17.5 %	62.0 %
Level 4	496	44.5 %	44.5 %
Grades 6–8			
Level 1	—	22.1 %	100.0 %
Level 2	447	18.5 %	77.9 %
Level 3	489	15.3 %	59.3 %
Level 4	514	44.0 %	44.0 %
Grade 10			
Level 1	—	25.3 %	100.0 %
Level 2	463	25.0 %	74.7 %
Level 3	506	16.1 %	49.7 %
Level 4	535	33.6 %	33.6 %

Exhibit 4.9: Percentage of Students at Each Performance Level—Social Studies

Performance Level	Scale Score Cut Score	Percentage in Level	Cumulative Percentage (at and above) for Each Performance Standard
Grades 3–5			
Level 1	—	19.3 %	100.0 %
Level 2	423	32.7 %	80.7 %
Level 3	492	30.1 %	48.1 %
Level 4	549	18.0 %	18.0 %
Grades 6–8			
Level 1	—	19.7 %	100.0 %
Level 2	439	27.3 %	80.3 %
Level 3	503	34.1 %	53.0 %
Level 4	560	19.0 %	19.0 %

Chapter 5: Technical Characteristics and Interpretation of Student Scores

This section describes the psychometric analyses conducted as part of the South Carolina Alternate Assessment (SC-Alt) 2007–2010 operational administrations. In 2010, only independent field-test tasks and items in biology were newly calibrated and evaluated. However, in order to provide a complete description of the technical characteristic of the 2010 assessment in all content areas, this chapter also reports the data analysis results for the sections of the assessment that had previously been calibrated using the 2007, 2008, and 2009 operational data (see American Institutes for Research and South Carolina Department of Education, 2008, 2009). The reported analyses are intended to ensure the quality of the items, the assessment materials and instruments, and the score reporting scales as measures of state academic standards.

As a reminder to the reader, there are three grade-band forms in each content area: elementary school (grades 3–5), middle school (grades 6–8), and high school (grade 10; except social studies). At each grade-band, the assessments have three potential starting tasks that correspond to three levels of task complexity (low, moderate, and high). Students are assigned to a starting task on the basis of teacher judgments recorded in the Student Placement Questionnaire (SPQ) for each content area. Linking tasks connect the grade-band forms so that the vertical test scale could be created.

Analysis and Scaling of Items, Tasks, and Test Forms

The ELA, mathematics, science, and social studies assessments underwent comprehensive psychometric analyses, including initial item calibrations, after their earlier field-testing. Final calibrations were estimated for the ELA, mathematics, and science content areas on the basis of operational data gathered during the spring 2007 operational administration; final calibrations for social studies were computed from operational data from the spring 2008 administration. Calibrations based on operational data were considered superior to those based on field-test data. The vertical scales were also defined using the linking tasks as the vehicle that connected the elementary, middle, and high school forms.

AIR calibrated the operational items, estimated examinee proficiencies, and calculated scale scores and achievement levels for operational forms. This process entailed examining item statistics to ensure quality measurement across the range of the assessment, calibrating the items within each content area to a common scale, and then applying a maximum-likelihood (ML) scoring algorithm to each student's responses to estimate his or her proficiency scores and assign the correct achievement level.

Assignment of Examinees to Starting Tasks and Item Calibration and Test Forms Linking

All eligible students participated in the spring 2010 test administrations. The sample sizes of approximately 1,400 students in elementary, 1,200 in middle school, and 400 in high school, per content area, enabled effective calibration across task starting points and grade-bands. Students were assigned to one of three starting points on the basis of the sum of the teacher responses on the SPQ. The SPQ cut scores were shown to correlate with student achievement scores on the 2006 field-test administrations (for details, see American Institutes for Research, 2008). The assignment of student starting tasks based on the SPQ cut scores was intended to expose students to items that were ideally suited to their current level of achievement while ensuring that (a) each student responded to an adequate number of items so that reliable and content-valid proficiency

scores could be estimated and (b) an adequate number of students responded to each item for the joint calibration to be reliable.

Teachers were instructed to administer all tasks associated with the assigned starting point, with provisions for dropping to a lower starting point (task) if the student was unable to respond to the items in the task at the assigned starting point. Students who were assigned to high and moderate levels of the assessment but were unable to respond to items in the tasks at those levels may have been moved back to a less difficult starting point.

The linking design allowed a joint (concurrent) calibration of all items within a content area and the placement of the items on a common difficulty scale. The tasks actually used to link the grade-band forms (linking tasks) were selected, in part, on the basis of their moderate difficulty levels. Moderately difficult tasks contribute to more stable linking across levels than tasks that may be either too easy or too difficult for the examinees.

Linking across grade-band forms was accomplished by using common tasks across grade-bands. Some of the tasks from the elementary form are on the middle school form; some of the tasks from the middle school form are on the high school form. (For details, see "Linking Tasks in Each Grade-Band Assessment" above.) In general, tasks are assigned in such a manner that the forms increase slightly in difficulty as examinees progress through the grade-bands. This means that a task assigned to the moderate level of complexity in the elementary form may be assigned to the low-moderate level in the middle school form.

See Appendix E for a summary of the linking design in each of the four content areas.

Analysis Plan

AIR's analyses presented in the remainder of this chapter were conducted in five steps:

1. Data preparation and quality control
2. Classical item analysis
3. Review of items not meeting psychometric criteria for inclusion on operational forms
4. Joint calibration of items according to the Rasch model
5. Final achievement estimation and scale score calculation for operational forms

Data Preparation and Quality Control

Before analyzing the operational test data, AIR psychometricians performed a number of quality control procedures to ensure that scanning operations resulted in accurate data capture of the teacher-recorded student responses. Prior to the test administration, AIR verified all of the point values for each form's answer folder. For each form, two AIR staff members independently verified the possible responses and point values for each item.

After receiving the scanned test data, AIR analysts carefully examined the data file to verify its accuracy. Descriptive statistics were computed to ensure that student case counts on the pre-identification file generally corresponded to the actual counts based on test data at the state, school, and classroom levels. In addition, AIR verified that the total number of items in the data

file matched the number of items on the answer folder and in the test booklet and then examined the frequency distributions of item responses to identify potential scoring problems, such as out-of-range values or unused response categories.

For purposes of item analysis and student scoring, respectively, non-response (NR) data were treated in two different ways:

For *item analysis and calibration purposes*, a student had to have at least three scored responses for the testing attempt to be considered valid. For a response to be considered a scored response, the test administrator had to have assigned a numeric score (0–4) to the student's response. If the administrator scored NR for all items in a task, the task was treated as not administered, and NR values were recoded as missing.

For *operational scoring* of student responses and estimation of student proficiency, however, the NR codes were treated as indications that the item was administered and that the student did not possess the content area knowledge and skill to respond. In this case, all NR values were recoded as zeroes and included in the student proficiency estimates. Following this recoding, tests were reexamined to determine the number of scored responses (0–4) in each content area. For operational scoring, a student had to have at least five scored responses of any kind for the assessment to be considered a valid attempt within a content area.

After the accuracy of the data file was verified, classical item analyses and Item Response Theory (IRT) analyses were performed. Several quality control procedures were taken to ensure the accuracy of these analyses.

As an initial step, the program control file was checked by two data analysts to ensure that form layout was correctly specified and that item response values were correct. As a second step, two analysts independently performed all analyses. Results of the parallel analyses were compared for mistakes by using commercially available file comparison software. Last, the analysis results were spot-checked by using other commercially available statistical software to ensure that the results were consistent across statistical software packages. *These comprehensive quality control steps are highly effective in detecting any issues that might influence the interpretation of the item analysis results.*

Classical Item Analysis

Classical item analysis for the SC-Alt operational and field-test forms was conducted using the *AM* statistical software (http://am.air.org). The item analysis yielded the proportion of students in each response category, the percentage of omitted responses for that item[2], and the proportion of students who were unable to respond to the item because of access limitations (where relevant). Correlations between the item score and the test score were computed using adjusted polyserial correlations. For purposes of calculating item statistics, omitted items were treated as incorrect when there was at least one scored response within the same task (see above). Minimum and maximum point values, average item scores, and adjusted item-total polyserial correlations were calculated for all items.

[2] An item was considered omitted if no response was recorded for the item (or the test administrator marked NR on the student score sheet), but the student responded to subsequent items on the task.

Test form statistics, such as internal consistency reliability estimates and standard error of measurement statistics, were suppressed at this point because all students were not expected to take all items. Such statistics would be misleading before Rasch scoring was applied. Special marginal reliability analyses used to determine the reliability of the student score estimates are described in a later section of this chapter.

The proportion of students in each score-point category was calculated as defined by the item's scoring guidelines, as well as the proportion of students with blank responses within attempted tasks (i.e., those with at least one scored response). Item difficulty was computed as the mean score on the item across all students taking the form and with a scored response on that item. The average proportion of total points, calculated as the mean score divided by the total number of points possible on the item, serves as an additional measure of item difficulty.

Review of Items Not Meeting the Specified Psychometric Criteria

Classical item analysis provided information about the technical quality of the items; items failing to meet specified psychometric criteria were flagged for subsequent review. During field-testing of ELA and mathematics (spring 2006), science (fall 2006), and social studies (spring 2007), AIR reviewed all flagged items in concert with SCDE to determine whether they were of sufficient psychometric quality. For the 2007 operational forms in ELA, mathematics, and science, and for the 2008 operational form in social studies, AIR conducted a statistical review of the items to determine whether any operational items were performing in an unacceptable fashion. For the spring 2009 operational SC-Alt administration, AIR subjected all embedded field-test items in ELA, mathematics, science, and social studies to an item data review. The spring 2010 operational SC-Alt administration contained only operational items in ELA, mathematics, science, and social studies. Items from the 2010 spring independent field test in high school biology were subjected to similar analyses as previous field-test items (see Chapter 9).

Item Response Theory Calibration and Linking Test Forms

This section describes AIR's procedures for item calibration using IRT techniques. Item parameters were estimated using the Partial Credit model (Masters, 1982) approach available using Winsteps software. A common item design was used to enable simultaneous calibration and linking across grade-band test forms in each content area. Items were jointly calibrated across grade-bands in a single Winsteps run for each content area. This calibration approach put the item parameters of all grade-band test forms within a content area on the same scale.

For 2010, the results reported on the vertical scale appear in Exhibit 5.1 and Exhibit 5.2. It is interesting to note that the mean scores show a general upward trend across grades. This indicates that a vertical scale is a useful way to describe the results of this population of students. The exceptions are the mean mathematics and science scale scores in high school, which are within one-fifth of a standard deviation from the middle school means. In addition, in almost every grade-band, a few students were at the floor of the test (minimum scaled score equal to 260), but very few reached the ceiling (maximum scale score equal to 740).

Exhibit 5.1: Scale Score Statistics, by Grade-Band, Overall

Subject	Statistic	Elementary School	Middle School	High School
ELA	N	1391	1184	357
	Mean	496.57	511.87	512.11
	SD	61.19	75.83	77.33
	Min	260	260	260
	Max	663	717	740
Math	N	1386	1184	355
	Mean	501.05	508.3	495.07
	SD	69.18	75.01	71.8
	Min	260	260	260
	Max	698	718	732
Science	N	949	809	356
	Mean	507.37	514.07	503.49
	SD	66.96	79.05	84.55
	Min	260	260	260
	Max	730	730	740
Social Studies	N	951	818	0
	Mean	503.3	519.89	.
	SD	76.9	90.56	.
	Min	260	260	
	Max	740	734	

Exhibit 5.2: Scale Score Statistics, by Grade-Band, by Primary Disability

		Elementary School				Middle School				High School			
Subject	**Statistic**	**Severe**	**Moderate**	**Mild**	**Autism**	**Severe**	**Moderate**	**Mild**	**Autism**	**Severe**	**Moderate**	**Mild**	**Autism**
ELA	N	146	338	387	321	135	409	274	234	38	161	64	57
	Mean	407.62	486.32	533.7	493.64	416.48	501.34	573.7	504.21	400.37	517.72	578.6	488.07
	SD	67.25	42.55	42.8	47.09	73.48	44.7	65.65	59.57	84.81	48.68	59.73	60.78
	Min	260	260	412	321	260	260	377	260	260	372	472	313
	Max	529	663	663	663	607	679	717	717	578	740	740	604
Math	N	147	338	385	318	135	410	274	233	38	160	64	56
	Mean	404.86	487.16	543.7	497.1	411.19	496.92	569.3	503.74	387.97	499.72	554.1	482.66
	SD	69.41	44.66	52.54	51.67	71.5	46.3	61.4	59.56	69.62	44.52	53.33	63.51
	Min	260	260	423	288	260	260	388	260	260	260	454	260
	Max	521	698	698	698	613	670	718	718	501	645	732	570
Science	N	104	219	277	217	86	274	196	170	38	162	64	55
	Mean	414.77	499.57	541.8	504.87	402.52	508.05	572.9	496.35	384.24	511.57	571.2	480.04
	SD	73.14	47.03	44.6	57.9	80.04	49.63	58.28	60.23	85.39	59.85	67.82	62.36
	Min	260	260	418	260	260	260	445	260	260	297	439	302
	Max	527	730	730	730	577	668	730	668	571	680	740	604
Social Studies	N	106	237	249	225	96	277	193	158	0	0	0	0
	Mean	390.12	496.47	551.1	499.57	400.66	512.66	595.5	503.63	.	.	.	.
	SD	77.84	56.2	50.07	63.45	86.5	54.7	69.94	64.57	.	.	.	.
	Min	260	260	435	260	260	260	345	312	.	.	.	.
	Max	561	740	740	740	669	734	734	734	.	.	.	.

The South Carolina alternate assessment implements vertical scaling, permitting the measurement of student progress on the state content standards over time. Such a scale can provide educators and parents with useful information for monitoring student performance as students move through grades over time.

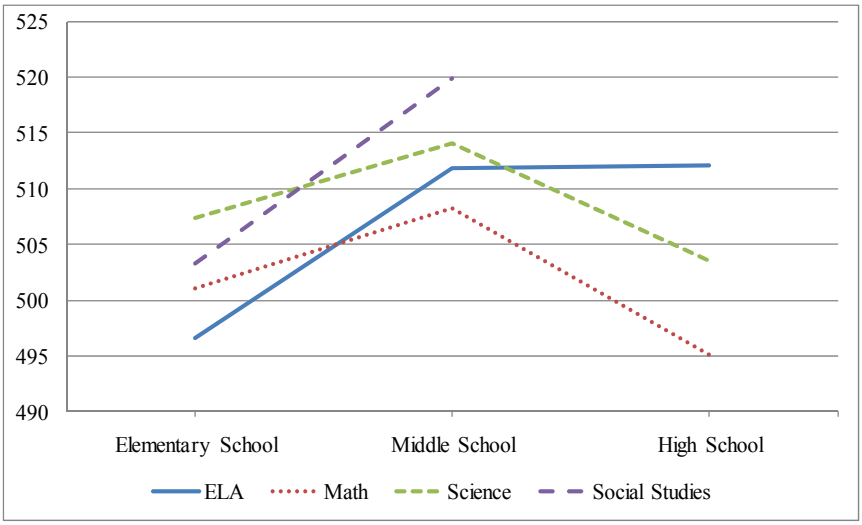
The development of this scale required the use of a common item linking design. In a common item design, *linking item tasks* appear on two adjacent grade-band forms of the assessment. These linking items allow for the grade-band scales to be connected, thus establishing the vertical scale.

There are at least two features of this linking design and the use of grade-level forms that warrant clarification. First, the linking items are the same (i.e., they are the same items) between two grade-band forms. They are connected to grade-specific standards in the higher grade as well as in the lower grade. As a result, students are not exposed to off-grade-level content since the common items serve a dual purpose in measuring content in both grade-bands. Second, even though some items on each grade-band form are administered for 2-3 years before replacement, it is not likely that exposure of the items to the students would trigger responses based on the recollection of any item's administration the previous year. As students grow academically, their starting task will likely change each year. New starting tasks mean that a portion of the items any student receives will be unique each year. New task development and field testing during 2009 allowed some new tasks and items to be added to the 2010 forms to replace older tasks. New tasks and items will be added to the 2011 and 2012 forms, replacing selected older tasks.

The linking design will change for future forms. For the 2011 forms, in the area of science, biology will be added as an operational assessment at the high school level without linking to the earlier grades. Vertical linking will also be discontinued for the elementary and middle school science forms. A linking design will be maintained in ELA and mathematics, however, beginning with the 2012 forms, the scores of ELA and mathematics linking tasks will contribute operationally to student performance scores at only one grade-band.

A graph of the overall pattern of performance for 2010 on the vertical scale is shown in Exhibit 5.3. Again, there is a general upward trend from elementary to middle school grade-bands in each of the four subject areas. This graph shows that the vertical scale in the SC-Alt was successful at capturing growth across grade-bands.

Exhibit 5.3: Overall Pattern of Performance on the Vertical Scale



Using Item Responses to Estimate Student Proficiency

This section describes the estimation of student proficiency for the SC-Alt operational administration of English language arts, mathematics, and science assessments for elementary, middle, and high school; social studies assessments for elementary and middle school are also reported. The section first describes the estimation procedures used to determine student proficiency based on the items administered, then the transformation of proficiency estimates on the Rasch theta scale into scale scores, and finally the relation of achievement estimation to reliability estimation.

Student proficiency scores were estimated using a maximum-likelihood approach based on the scored items for each student.[3] This method calculates the theta score that maximizes the likelihood function of the given item responses for each student. Comparable scale-score estimates from these different item responses were achieved through the measurement-invariance property of IRT ability estimates, even when students were exposed to different ranges of items.

Under the Rasch-based IRT model, there is a one-to-one correspondence between the estimated theta score and the total raw score for a specific set of items. However, in the SC-Alt, each student can take different sets of items. Using the pattern scoring method for calculating theta scores, we ensured that (a) two students who took the same items and achieved the same item scores were assigned the same theta score, and (b) students who took more difficult items were assigned higher theta scores than students with the same raw scores who took less difficult items. Thus, the scoring method took into account both the number of raw score points the student achieved and the difficulties of the items the student responded to. This scoring process was performed separately for each content area.

Once theta values had been estimated for each student, AIR converted the theta estimates to scale scores using a scale metric determined by SCDE in consultation with AIR. The SC-Alt

[3] The first step in this process was to rescore student responses consistent with the operational scoring method described under the "Data Preparation and Quality Check" section.

assessments were scaled to have a mean of 500 and a standard deviation of 80 on the vertical scale for the grade-band 6–8 assessment. The grade-band 3–5 and grade 10 assessment means and standard deviations were calculated in relation to grade-band 6–8 mean and standard deviation. This was done by performing a linear transformation of the Rasch theta scale for each content area, fixing the mean of the middle school test form scale at 500, and multiplying the student's theta deviation score by 80 as shown in the formula below:

$$y_{ijk}^{*} = 500 + \left(\frac{\hat{\theta}_{ijk} - \hat{\mu}_{k}}{\hat{\sigma}_{k}}\right) * 80 ,$$

where

i indexes student;

j indexes grade-band;

k indexes content area;

y_{ijk}^{*} is the scale score for student *i* in grade-band *j* and content area *k*, given estimated ability, θ_{ijk};

$\hat{\mu}_{k}$ is the content-area-specific mean for the middle school test form; and

$\hat{\sigma}_{k}$ is the content-area-specific standard deviation for the middle school test form.

SCDE also decided to truncate the scale score ranges so that the lowest possible scale score was 260 and the highest possible scale score was 740. Student scale score estimates were truncated to the smallest whole integer (e.g., an estimated scale score of 440.60 would become 440). Additionally, scale scores were calculated and checked using a method similar to the process for total raw data.

Once scoring was completed, it was possible to estimate the internal consistency score reliability of the grade-band assessments by estimating the marginal measurement error across students. These estimates produced different standard errors for each student, depending on the items they were given and their level of performance on those items. This value was used to determine the score reliability as the proportion of true score variance to observed score variance. We estimated this value within each content area (a) across the entire theta scale, (b) across grade-band forms, and (c) for each starting point within a grade-band.

Test Score Reliability

This section provides the marginal reliability for each grade-band, content area, and groups of students beginning at each starting task determined by the SPQ for the spring 2010 administration.

Classical test theory-based reliability indices, such as Cronbach's alpha, were not appropriate for the SC-Alt because the length of the test and the subset of items differed for each student. The reliability coefficient for the SC-Alt was, therefore, calculated as the *marginal reliability* (Sireci,

Thissen, & Wainer, 1991), which is equivalent in interpretation to classical internal consistency estimates of reliability.

First we determined the marginal measurement error variance, $\overline{\sigma}_{e^*}^2$, across all examinees:

$$\overline{\sigma}_{e^*}^2 = \int \sigma_{e^*}^2 p(\theta) d\theta = \frac{\sum \sigma_{e^*}^2}{N},$$

where $\sigma_{e^*}^2$ is the square of the standard error of student ability estimate, $\hat{\theta}$. Thus, the marginal measurement error variance could be estimated as the average of squared standard error of $\hat{\theta}$.

Then we estimated the marginal reliability as

$$\overline{\rho} = \frac{\hat{\sigma}_{\theta}^2 - \overline{\sigma}_{e^*}^2}{\hat{\sigma}_{\theta}^2},$$

where $\hat{\sigma}_{\theta}^2$ is the variance of observed θ estimates.

The marginal reliability estimate, $\overline{\rho}$, can be interpreted similarly to classical reliability indices such as Cronbach's alpha. Estimates of the marginal reliability for the test forms corresponding to the three SC-Alt grade-band assessments can be seen in Exhibit 5.4.

Exhibit 5.4: Marginal Reliability and Standard Error of Measurement by Grade-Band and Subject

Subject	Grade-Band	Elementary	Middle	High
English Language Arts	N	1391	1184	357
	Reliability	0.920	0.897	0.920
	$\overline{\sigma}_{e^*}$	17.306	24.376	21.929
Math	N	1386	1184	355
	Reliability	0.901	0.912	0.919
	$\overline{\sigma}_{e^*}$	21.766	22.234	20.460
Science	N	949	809	356
	Reliability	0.872	0.891	0.891
	$\overline{\sigma}_{e^*}$	23.962	26.139	27.934
Social Studies	N	951	818	x
	Reliability	0.906	0.877	x
	$\overline{\sigma}_{e^*}$	23.556	31.751	x

The marginal reliability estimates for ELA and mathematics met or exceeded 0.90 in each grade-band form; the marginal reliability estimates for science met or exceeded 0.87; and those for social studies met or exceeded 0.88. The reliability estimates of all four content areas fall into the range of reliability coefficients found with large-scale assessments (Rudner & Schafer, 2001) and meet the reliability requirements for assessments used for the purposes for which the SC-Alt was designed.

In addition to the marginal reliability estimates, Exhibit 5.4 also displays the marginal standard errors of measurement for each subject and grade-band, labeled $\bar{\sigma}_{\varepsilon^*}$. These marginal standard errors of measurement range between 17 and 32 scale score units, placing the SEM at approximately one-third of a standard deviation of the content area and grade-band. Appendix G shows the marginal reliability estimates broken out by groups of students beginning at each starting task, as determined by the SPQ. The reliability coefficients in Appendix G are generally somewhat attenuated compared with those in Exhibit 5.4, due to the reduction in variance of scale scores grouped by starting task.

Classification Accuracy

This section describes the extent to which student achievement-level classifications were accurate across students. Classification accuracy was estimated for each cut score as the average probability of correct achievement-level assignments across all examinees (assignments above or below the cut score), given each examinee's estimated proficiency score, θ_i:

$$\mathrm{CA_K} = \frac{\sum_{i=1}^{N_{k\geq K}} P(\theta_i > \theta_K^* \mid \theta_i, k_i \geq K) + \sum_{i=1}^{N_{k<K}} [1 - P(\theta_i > \theta_K^* \mid \theta_i, k_i < K)]}{N},$$

where

θ_i is the proficiency (i.e., theta) of student *i*;

k_i is the assigned performance level of student *i*;

θ_K^* is the cut score for the performance level *K* on the theta scale; and

N is the sum of the number of students at or above the cut score, $N_{k\geq K}$, and the number of students below the cut score, $N_{k<K}$, or simply the total number of students.

Thus, $P(\theta_i > \theta_K^* \mid \theta_i, k_i \geq K)$ is the probability that a student with θ_i assigned to achievement level k_i is above the cut score, θ_K^*. The classification accuracy is the expected rate of correct classification probability, ranging from 0 to 1, where higher values indicate superior classification consistency. Exhibit 5.5 shows the classification accuracy by content areas, achievement levels, and grade-bands.

Exhibit 5.5: Classification Accuracy

Subject	Achievement Level	Elementary School	Middle School	High School	Overall
English Language Arts	Level 2	0.991	0.989	0.982	0.989
	Level 3	0.936	0.938	0.949	0.938
	Level 4	0.944	0.951	0.939	0.946
Mathematics	Level 2	0.983	0.973	0.980	0.979
	Level 3	0.930	0.942	0.927	0.935
	Level 4	0.941	0.943	0.947	0.943
Science	Level 2	0.982	0.976	0.963	0.976
	Level 3	0.944	0.935	0.921	0.936
	Level 4	0.907	0.933	0.914	0.918
Social Studies	Level 2	0.980	0.976	x	0.978
	Level 3	0.932	0.946	x	0.938
	Level 4	0.932	0.925	x	0.929

For example, according to the estimates in Exhibit 5.5, for the grade-band 3–5 English language arts assessment, 99% of students were correctly classified at Level 1 and 94% of students at Level 2 or above. All students in all grade-bands had a probability greater than 0.93 of being classified accurately as proficient (i.e., as level 3 or higher).

These results indicate that the measurement errors at the performance level cut points (Exhibits 4.1–4.4) are small compared to the overall variance of student performance.

The calculation of probability of the correct performance level for students is described in the following section.

Calculation of the Probability of Being Classified Above a Cut Score Given the Student's Theta Score

For each student we can compute the likelihood of theta $L(\theta \mid \mathbf{z}, \mathbf{b})$. Suppose that the prior of the theta distribution is $f(\theta)$. Then, using Bayes' rule, we have

$$f(\theta \mid \mathbf{z}, \mathbf{b}) \propto f(\theta) L(\theta \mid \mathbf{z}, \mathbf{b}),$$

where $L(\theta \mid \mathbf{z}, \mathbf{b})$ is the likelihood of theta given the response $\mathbf{z}$ and item parameters $\mathbf{b}$; hence, the probability at and above cut is

$$P = \frac{\int_{\theta \geq \theta_{\text{cut}}} f(\theta) L(\theta \mid \mathbf{z}, \mathbf{b}) d\theta}{\int f(\theta) L(\theta \mid \mathbf{z}, \mathbf{b}) d\theta},$$

where $f(\theta)$ can take different distribution such as normal, or uniform, depending on our prior belief.

Calculation of $L(\theta | \mathbf{z}, \mathbf{b})$

For the Rasch model, we have

$$L(\theta \mid \mathbf{z}, \mathbf{b}) = \prod_{i \in MC} \left(\frac{Exp(z_i\theta - b_i z_i)}{1 + Exp(\theta - b_i)} \right) \prod_{i \in CR} \left(\frac{Exp(z_i\theta - \sum_{k=1}^{z_i} b_k)}{1 + \sum_{i=1}^{K_i} Exp(\sum_{k=1}^{i} (\theta - b_k))} \right)$$

$$\propto Exp(r\,\theta) \prod_{i \in MC} \left(\frac{1}{1 + Exp(\theta - b_i)} \right) \prod_{i \in CR} \left(\frac{1}{1 + \sum_{i=1}^{K_i} Exp(\sum_{k=1}^{i} (\theta - b_k))} \right)$$

where K_i is the maximum score for item *i* when this item is a CR item. It can be noted that the calculation above depends on total raw score r only when using the attempted items.

Chapter 6: Score Reports

This chapter describes the method used for reporting scores on the SC-Alt for the spring 2010 administration. An Individual Score Report (ISR) is included in Appendix H as an example of the highly detailed and diagnostic nature of the reports. This chapter gives a brief overview of how scores on the SC-Alt are reported; a more detailed description is available in a separate *Score Reports User's Guide*.

The SC-Alt has three types of score reports: the ISR, or family report; school reports; and district reports. Each report conveys specific information to its target audience. The reports are designed to be easily used by parents and educators. Of particular note, the reports include in-depth information about what students know and can do relative to the South Carolina academic content standards and to the performance levels.

The ISR provides specific performance feedback for each student across four content areas: mathematics, English language arts, science, and social studies (added in 2008). Within each content area, a graphic bar highlights the student's performance level along the proficiency scale. Each performance level is described in broad, easy-to-understand content terms. Further descriptions of what a student knows and can do are tailored and printed for each obtained performance level. For example, if a student is classified as Level 3 in mathematics, the following message is printed: "Students who score at Level 3 should be able to add and subtract simple numbers, count and compare objects in a group, compare objects by color, size, or shape, identify three-dimensional shapes, and read information in a graph." (Note: Scale scores were added to the ISR starting with the spring 2008 reporting cycle.)

Specific activities, based on each student's performance level for each content area, are presented for the family to do at home to help ensure positive academic growth in the content area.

The school report provides a summary of the performance of each student in the school. The alphabetical list of students contains basic demographic information and test form administered, in addition to achievement data. A scale score and achievement level are listed for each student for each content area. A school summary shows the number of students scoring at each performance level.

Three district-level reports are issued. The district roster summary report displays the roster of the district's tested students along with their demographic information, their scale scores and performance levels for each content area, and type of test form. The district summary by test form report presents a roster of schools in which students were tested, identifying the test form and giving the number of students tested in each content area and the percentages achieved in each performance level by content area. The total number of students tested with each form and their performance-level distributions by content area are listed at the bottom of the report. The district demographic summary report shows the number of students tested and the distributions of performance levels in all content areas, disaggregated by gender, ethnicity, lunch program, migrant status, and ESL status.

The separate *Score Reports User's Guide* has more specific information on how to interpret student scores and score reports and how to relate academic growth as measured by the SC-Alt to classroom curricula and activities. The guide has been widely distributed throughout South Carolina.

Chapter 7: Student Performance Data from the Spring 2010 Administration

Performance data from the spring 2010 administration are presented in this chapter. This was the fourth operational administration of the SC-Alt ELA, mathematics, and science assessments and the third operational administration of the SC-Alt social studies assessment.

A total of 2,941 students from 83 school districts and 527 schools were tested with the SC-Alt in spring 2010. The total number of tested students with one or more valid content area scores was 1,393 for the elementary form, 1,185 for the middle school form, and 357 for the high school form. Five of those students tested on two forms each, and another six students did not have valid responses in any of the content areas.

Nearly one-half of the participating school districts (35; 42%) tested 15 or fewer students; 33 districts (40%) tested 16 to 50 students; and 15 districts (18%) tested more than 50 students each. Eight districts tested more than 100 students; the greatest number of students tested in one district was 250.

Of the 527 schools testing SC-Alt students, 342 (65%) tested five or fewer students; 115 (22%) tested six to 10 students; 60 (11%) tested 11 to 20 students; and 10 schools (2%) tested 21 or more. Only two schools tested more than 50 students each (64 and 69 students).

The elementary school form was developed to be administered to students who are 8, 9, or 10 years old at the beginning of the school year, which are the ages typical of students enrolled in grades 3, 4, and 5. The middle school form was developed for students who are 11, 12, and 13 (typical of students enrolled in grades 6, 7, and 8), and the high school form was developed for students age 15 (typical age of students in grade 10).

Students tested with the elementary and middle school forms with reported ages outside the specified age ranges were either erroneously assigned to the forms by the test administrator or, in some cases, took the test as a result of birth date coding errors on the data files. The number of students reported outside the expected ages for the elementary and middle school forms is less than 1% for each content area. Students reported as having been tested on the high school form with ages below 15 were tested because of errors in form assignment or birth date coding. Students older than 15 (e.g., 16) may be assessed with the high school form if they have not been assessed at the high school level previously.

The performance of students by grade-band form, age, and demographic group for the ELA, mathematics, science, and social studies content areas is presented in Appendix I.

Chapter 8: Validity

Content Validity

One source of evidence for the content validity of the South Carolina Alternate Assessment was obtained through independent alignment studies. The University of North Carolina at Charlotte (UNCC) conducted studies of the alignment of (a) Assessment Standards and Measurement Guidelines (ASMGs) to grade-level curriculum standards and (b) SC-Alt items to the ASMGs that they targeted. This was a pilot study conducted by Flowers, Browder, Wakeman, and Karvonen with UNCC through the National Alternate Assessment Center (NAAC). (South Carolina is a member state of the NAAC.) A second independent study of ELA and mathematics was completed by the South Carolina Education Oversight Committee (EOC; 2008a) as required by the state Education Accountability Act of 1998 (EAA). The EOC approved the ELA and mathematics content areas on February 28, 2008. The UNCC alignment study results for the English language arts and mathematics assessments are reported in detail in Flowers, Browder, Wakeman, and Karvonen (2006a). The results of the alignment studies for the ELA and mathematics assessments indicate that

> The state has evidence supporting alignment for its measurement guidelines and alternate assessment based on all seven criteria. We conclude that overall this is an alternate assessment system that links to the grade level content. Some areas for consideration in further development of the system are noted related to balance of content. (p. 7)

The alignment study results for the science assessment are reported in detail in Flowers, Browder, Wakeman, and Karvonen (2006b) and in an addendum dated December 21, 2007. The results of the alignment study for the science assessment indicate that

> The strength of the South Carolina science Alternate Assessment was that nearly all of the content was academic science content (98%). This is especially notable given that the alternate assessment tasks included items accessible to students at all symbolic levels. In contrast, the degree of alignment of AA tasks/items to grade-level standards was lower than those found in the alignment of ELA and mathematics. This difference could be due to the fact that the state's science grade-level standards changed during the development of the science AA. Another challenge was that the state had linked its alternate assessment tasks to the state standards and not directly to the measurement guidelines, creating a tough challenge to demonstrating alignment…. Our work with other states suggests that science may typically be the area rated as having the weakest alignment. (p. 4)

SCDE reviewed the initial science alignment study and determined that one source of some misalignment had resulted from the linking of some items to multiple standards and indicators in the alignment document provided by SCDE. During the Science Content Review Committee meeting, some members recommended adding additional indicators to align to some items. The intent of these recommendations focused more on instruction and demonstrating that instruction could include multiple standards and indicators. However, the alignment study team considered only the first two standards aligned to each item. In some cases, the first two standards were not necessarily the most appropriate. SCDE prioritized the standards and indicators and resubmitted the documentation for an additional study. From this review, completed December 21, 2007

(Flowers, Browder, Wakeman, & Karvonen, 2007), 163 of 173 items were rated as academic. Of the 10 items listed as nonacademic, six were rated as foundational (p.1). SCDE is currently addressing the items that were rated as having no content centrality by developing replacement items for new forms.

At the time of the alignment study for ELA and mathematics by Dr. Flowers and colleagues, the design of the SC-Alt was envisioned as a single assessment across grade levels. This design changed to a grade-band assessment following the study; however, the information provided from the alignment study was used to identify items with alignment difficulty, and these items were omitted from the operational grade-band test forms. Information from the review along with teacher comments was also used during item data review as part of the decision-making process regarding inclusion of items in the assessment.

A second independent review of the alignment of the science assessment was conducted by the Education Oversight Committee (EOC; 2008b). The EOC approved the elementary and middle school science alternate assessment on August 12, 2008. The EOC alignment findings were based on the review of two sets of studies of the SC-Alt:

- Studies of the alignment between the SC-Alt science assessment and the state academic standards conducted by University of North Carolina-Charlotte and Western Carolina University professors of curriculum and special education, in cooperation with the South Carolina State Department of Education (SCDE) and the National Alternate Assessment Center (Flowers, Browder, Wakeman, & Karvonen, 2006a, 2006b, 2007)
- A technical review of the task and item data from the 2007 test administration conducted by a professor of educational research and assessment at the University of South Carolina

Copies of the reports of the EOC reviews and findings are available in their entirety from the SCDE. Based on this review, the EOC identified a number of strengths of the SC-Alt science assessment that were noted in the final report:

- The assessment provides accountability and information for instructional improvement for students with significant cognitive disabilities who would not otherwise be assessed in the state testing programs, even with test accommodations and modifications.
- The assessment is intended to be aligned with the same grade-level academic standards as for all students, although at levels of complexity appropriate for the diversity of cognitive functioning observed among students with significant cognitive disabilities.
- The assessment format allows each student to respond to the items using the communication modes the student uses during instruction, such as oral response, pointing, eye gaze, a response card, sign language, or an augmentative communication device.
- The procedures for placing the student at the appropriate level for beginning each assessment reduces student fatigue and maximizes the student's opportunities to show his or her highest performance;
- The items in the assessment have a wide range of difficulty, and the test is moderately able to discriminate between high and low levels of performance.

The EOC report noted that while 96% of the items were found to be aligned to science inquiry standard indicators, the alignment of the items to content standards was 78%, falling short of an expectation for successful alignment of 90% set by the original evaluators. The EOC recommended that the SCDE review the alignment of the SC-Alt science items to the grade-level standards and identify items needing revision or replacement.

The SCDE and its contractor, the American Institutes for Research (AIR), reviewed the alignment and the ASMGs and established priorities for development of tasks to fill identified gaps. During 2008, SCDE and AIR developed five new tasks consisting of 32 items to be used to replace poorly aligned items and improve content coverage in science. Three tasks were developed for the elementary science form, and two tasks were developed for the middle school form based on the findings of the alignment study. No new tasks were developed for the high school physical science test because a stand-alone biology field test was planned for spring 2010. In the future, biology end-of-course tests will replace the physical science end-of-course tests for purposes of reporting Elementary and Secondary Education Act (ESEA) participation calculations.

An independent review of the alignment of the new items by the Center for Research on Education (2009a) found that 98% of the new items were aligned to grade-level content standard indicators. Copies of the report of the alignment reviews and findings are available in their entirety from the SCDE.

A follow-up alignment study of the high school ELA and mathematics assessments and biology field-test items was conducted by the Center for Research on Education in October 2009, using the same procedures that were used for the elementary and middle school alignment studies in December 2006 and January 2007. Almost all (94% to 96%) of the items were rated as academic. This percentage exceeds the value typically found in alternate assessments (90%) according to the reviewers. The alignment study results are reported in detail in High School Alternate Assessment Alignment Report to the South Carolina State Department of Education (Center for Research on Education, 2009b).

Convergent and Discriminant Validity

According to Critical Element 4.1(e) of the federal peer review and Standard 1.14 of the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 1999), it is desirable, if not necessary, to provide evidence of convergent and discriminant validity. One common method for examining this aspect of validity is with a multitrait-multimethod matrix (MTMM) (Campbell & Fiske, 1959).

Campbell and Fiske (1959) proposed the MTMM matrix design as a tool for the study of convergent and discriminant validity in psychological measurement. The MTMM matrix employs a crossed factorial measurement design of traits and methods to reveal these types of validity in comparison:

- Large correlations on validity diagonals (i.e., same trait and different methods) indicate convergent validity.
- Low correlations in the heterotrait-monomethod blocks indicate discriminant validity and the absence of method effects.

- Low correlations in the heterotrait-heteromethod blocks also indicate discriminant validity.

Selection of Traits and Methods

The student's abilities in each of the subjects—ELA, mathematics, science, and social studies—make up the four traits for the MTMM study. Two methods are considered for assessing these traits: the Student Placement Questionnaire (SPQ), as a structured teacher rating of student's attainment, and the SC-Alt scale score, as an IRT-based indicator of the student's performance in each subject or trait. In other words, the two methods contrast test scores of student performance with expert (or teacher) ratings. With four traits and two methods, the MTMM correlation matrix is of order 8. Note that the high school assessment does not include a social studies component; therefore, the MTMM matrix for high school has only six rows and columns.

Results

MTMM matrices were computed separately for each grade-band. The results are given in Exhibits 8.1–8.3. Pearson correlations are used, with pairwise deletion of missing data. For each matrix, the minimum pairwise sample size is indicated. *P*-values of individual correlation coefficients are not reported since *all* correlations are significant ($p < 0.05$).

Exhibit 8.1: MTMM Matrix, Scale Scores with SPQ Scores, Elementary School

		IRT Scale Scores				SPQ Scores			
	Subject	**ELA**	**Math**	**Science**	**Social Studies**	**ELA**	**Math**	**Science**	**Social Studies**
IRT Scale Scores	**ELA**	1.000							
	Math	*0.896*	1.000						
	Science	*0.880*	*0.859*	1.000					
	Social Studies	*0.907*	*0.896*	*0.892*	1.000				
SPQ Scores	**ELA**	**0.771**	0.770	0.690	0.760	1.000			
	Math	0.770	**0.788**	0.702	0.755	*0.924*	1.000		
	Science	0.738	0.753	**0.685**	0.726	*0.884*	*0.911*	1.000	
	Social Studies	0.759	0.756	0.689	**0.753**	*0.910*	*0.903*	*0.922*	1.000
Minimum pairwise N: 459									

Exhibit 8.2: MTMM Matrix, Scale Scores and SPQ Scores, Middle School

		IRT Scale Scores				SPQ Scores			
	Subject	**ELA**	**Math**	**Science**	**Social Studies**	**ELA**	**Math**	**Science**	**Social Studies**
IRT Scale Scores	**ELA**	1.000							
	Math	*0.898*	1.000						
	Science	*0.902*	*0.910*	1.000					
	Social Studies	*0.895*	*0.896*	*0.926*	1.000				
SPQ Scores	**ELA**	**0.756**	0.780	0.773	0.777	1.000			
	Math	0.772	**0.808**	0.784	0.776	*0.931*	1.000		
	Science	0.765	0.775	**0.772**	0.791	*0.885*	*0.921*	1.000	
	Social Studies	0.759	0.791	0.799	**0.779**	*0.915*	*0.925*	*0.936*	1.000
Minimum pairwise N: 415									

Exhibit 8.3: MTMM Matrix, Scales Scores and SPQ Scores, High School

		IRT Scale Scores			SPQ Scores		
	Subject	**ELA**	**Math**	**Science**	**ELA**	**Math**	**Science**
IRT Scale Scores	**ELA**	1.000					
	Math	*0.892*	1.000				
	Science	*0.899*	*0.908*	1.000			
SPQ Scores	**ELA**	**0.778**	0.744	0.751	1.000		
	Math	0.775	**0.755**	0.752	*0.921*	1.000	
	Science	0.772	0.744	**0.747**	*0.886*	*0.918*	1.000
Minimum pairwise N: 315							

In each MTMM table, the *convergent validity coefficients* (correlations between measurements of the same trait using different methods) are marked in bold. These convergent validity coefficients range from 0.69 to 0.81 and certainly fall into an acceptable range. These high correlations demonstrate evidence for the validity of the SPQ; the three exhibits indicate that the SPQ and the actual test are essentially measuring the same trait and that the SPQ is a good indicator of performance on the test.

The *heterotrait-monomethod coefficients* in the monomethod triangles (correlations between measurements of different traits using the same method) are set in italics. These correlation coefficients range between 0.86 and 0.93 for IRT scale scores and between 0.88 and 0.94 for SPQ scores. The high overall range of these correlations indicates the presence of method variance. However, this is to be expected because the SPQ was not developed to measure the trait; instead, it only indicates the starting task on the test for measuring the trait. Such a result of high correlations in the monomethod triangles is not uncommon in MTMM studies (Fiske, 1995), and specific conditions offer themselves as causes for the present scenario. First, the different scale types—number-correct rating scales versus IRT scales of behavioral tests—are in

themselves a source of method variation; second, the SPQ's "can do" questions draw on the teacher's memory of a student's possible performance over the long term and are apt to differ in quality and veracity; and third, the IRT scale scores for the three subjects reflect the student's performance in the testing situation and are subject to the student's condition on the testing day.

The *heterotrait-heteromethod coefficients* appear in the tables in regular type. These correlation coefficients fall in the same range as the convergent validity coefficients, with values from 0.69 to 0.80. To confirm discriminant validity, the heterotrait-heteromethod correlations should be smaller than the convergent validity coefficients. Instead, these MTMM matrices support the notion that the four traits vary essentially along just a single dimension. Because the population of alternate assessment students is so *very* heterogeneous, the students' general level of cognitive functioning dominates the relationship among their scale scores.

Validity of the Student Placement Questionnaire (SPQ)

AIR reviewed item data from the 2010 administration regarding the agreement between SPQ recommended start points and the final observed start points. The purpose of the study was to determine the effectiveness of the SPQ in identifying the most appropriate starting task.

Administration of the SC-Alt uses the SPQ as a pre-assessment instrument to determine the most appropriate starting point in the assessment. The SPQ requires the teacher to evaluate the student on 12 to 15 "can do" statements addressing the student's skills and knowledge in each content area on the basis of the teacher's prior instructional knowledge of the student. A total score computed from the teacher's SPQ responses indicates the initial starting task for the assessment.

The instructions for using the SPQ require teachers to adjust the starting point below the SPQ recommended start point when the student is not successful on the first administered task. Alternately, after reviewing the assessment, some teachers may have judged that a student needed to start at a higher level than recommended by the SPQ.

A summary of the results of the agreement between the SPQ recommended start points and the observed start points for each content area and grade-band form is presented in Exhibits 8.4–8.7. These results indicate that the agreement between the SPQ recommended start point and the observed start point was 97% for ELA, 97% for mathematics, 96% for science, and 96% for social studies administrations. Use of the SPQ pre-assessment score is only the first step in the procedure used by the test administrator in determining where the student should start the assessment. Since the test administrator is required to make adjustments based on the student's success on the first task, and these adjustments are reflected in the agreement rates, the SPQ appears to be working very effectively for targeting the first task to begin the assessment process.

Exhibit 8.4: Agreement Between SPQ and Observed Start Points by SPQ Recommended Starting Tasks—ELA

Observed Start Task	Elementary School				Middle School				High School				Overall
	Recommended Starting Task												
	1	3	6	Total	1	3	6	Total	1	3	6	Total	
Starting task consistent with SPQ	97%	95%	98%	97%	95%	96%	99%	97%	95%	93%	98%	96%	97%
Lower start task than recommended	0%	4%	2%	2%	0%	3%	1%	1%	0%	6%	1%	2%	2%
Higher start task than recommended	1%	1%	0%	1%	2%	1%	0%	1%	2%	2%	0%	1%	1%
Nonstandard start task	1%	1%	0%	0%	1%	0%	0%	0%	1%	0%	0%	1%	0%
No valid test items; no starting task	1%	0%	0%	0%	1%	0%	0%	0%	2%	0%	0%	1%	0%
Incomplete SPQ	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Inconsistent with SPQ	3%	5%	2%	3%	5%	4%	1%	3%	5%	7%	2%	4%	3%
ELA Total (N)	348	334	709	1391	269	231	684	1184	98	54	205	357	2932

Exhibit 8.5: Agreement Between SPQ and Observed Start Points by SPQ Recommended Starting Tasks—Math

Observed Start Task	Elementary School Recommended Starting Task				Middle School Recommended Starting Task				High School Recommended Starting Task				Overall
	1	3	6	Total	1	3	6	Total	1	3	6	Total	
Starting task consistent with SPQ	96%	96%	98%	97%	95%	95%	99%	97%	97%	87%	97%	95%	97%
Lower start task than recommended	0%	3%	2%	2%	0%	3%	1%	1%	0%	10%	2%	3%	2%
Higher start task than recommended	2%	0%	0%	0%	4%	1%	0%	1%	0%	3%	0%	1%	1%
Nonstandard start task	1%	0%	0%	0%	0%	1%	0%	0%	1%	0%	1%	1%	0%
No valid test items; no starting task	1%	0%	0%	0%	2%	0%	0%	0%	2%	0%	0%	1%	0%
Incomplete SPQ	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Inconsistent with SPQ	4%	4%	2%	3%	5%	5%	1%	3%	3%	13%	3%	5%	3%
ELA Total (N)	358	358	670	1386	254	273	657	1184	90	68	197	355	2925

Exhibit 8.6: Agreement Between SPQ and Observed Start Points by SPQ Recommended Starting Tasks—Science

Observed Start Task	Elementary School				Middle School				High School				Overall
	Recommended Starting Task												
	1	3	6	Total	1	3	6	Total	1	3	6	Total	
Starting task consistent with SPQ	95%	98%	92%	95%	93%	98%	99%	97%	94%	92%	96%	95%	96%
Lower start task than recommended	0%	2%	7%	4%	0%	1%	1%	1%	0%	5%	3%	2%	2%
Higher start task than recommended	4%	0%	0%	1%	4%	1%	0%	1%	2%	2%	0%	1%	1%
Nonstandard start task	0%	0%	0%	0%	1%	0%	0%	0%	3%	2%	1%	2%	1%
No valid test items; no starting task	2%	0%	0%	1%	2%	0%	0%	0%	2%	0%	0%	1%	1%
Incomplete SPQ	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%
Inconsistent with SPQ	5%	2%	8%	5%	7%	2%	1%	3%	6%	8%	4%	5%	4%
Science Total (N)	305	235	409	949	206	165	438	809	109	61	186	356	2114

Exhibit 8.7: Agreement Between SPQ and Observed Start Points by SPQ Recommended Starting Tasks€Social Studies

Observed Start Task	Elementary School Recommended Starting Task				Middle School Recommended Starting Task				High School Recommended Starting Task				Overall
	1	3	6	Total	1	3	6	Total	1	3	6	Total	
Starting task consistent with SPQ	94%	95%	97%	96%	93%	95%	98%	96%					96%
Lower start task than recom-mended	0%	3%	3%	2%	0%	4%	2%	2%					2%
Higher start task than recom-mended	3%	2%	0%	1%	5%	0%	0%	1%					1%
Nonstandard start task	1%	1%	0%	1%	1%	1%	0%	0%					1%
No valid test items; no starting task	1%	0%	0%	0%	2%	0%	0%	0%					0%
Incomplete SPQ	0%	0%	0%	0%	0%	0%	0%	0%					0%
Inconsistent with SPQ	6%	5%	3%	4%	7%	5%	2%	4%					4%
Social Studies Total (N)	236	191	524	951	187	124	507	818					1769

Start-Stop Analysis

Data from the 2010 SC-Alt assessment were analyzed to address two questions concerning SC-Alt administration procedures and student performance:

1. How many tasks and items were administered to students who were started in the assessment at each of the three start points?
2. What was the achievement level performance of students who were started in the assessment at each of the three start points?

To address these questions, the task start point was identified for each student assessed by the 2010 administration of the SC-Alt assessment for all content areas and grade-band forms. According to each task start point, the number of tasks and items administered and the achievement level distribution were calculated and summarized.

SC-Alt test administrators were instructed to follow specific procedures concerning the use of the Student Placement Questionnaire (SPQ) to determine task start points, the minimum number of tasks to be administered, and whether to continue the administration through additional tasks until the student is no longer able to respond successfully. These procedures are detailed in Appendix B and in the 2010 *Test Administration Manual*, Appendix N.

Number of Tasks Administered

The minimum number of tasks to be administered is five tasks, when the test administration is started at Task 1, or seven tasks, when the administration begins at either Task 3 or Task 6. This rule is the same for every subject and form level. The actual number of tasks administered to students in the ELA, mathematics, science, and social studies content areas for each form level and task start point are presented in Exhibits 8.8–8.11.

In general, most students were administered at least the minimum number of tasks; the distribution of actual tasks administered often exceeded the minimum required when students were started at Task 1 or Task 3. In ELA, 54%, 60%, and 47% of students were administered more than five tasks when started at Task 1 in elementary, middle, and high school, respectively; 78%, 80%, and 81% were administered more than seven tasks when started at Task 3. Similar patterns are seen in math, science, and social studies, with students starting at Task 3 showing large percentages going beyond the minimum number of tasks administered.

Overall, fewer than 5% of students across forms and subjects were not administered the minimum number of tasks required. Among elementary and middle school administrations this rate ranges between 0.7% and 1.5%. In high school, however, the percentage of students administered fewer than the required number of tasks when the task start point was Task 1 was higher than for elementary or middle school. In high school, of students starting at Task 1, 5.9% in ELA, 7.1% in math, and 4.5% in science were administered fewer than five tasks.

Students whose started at task 1 were administered between 5.5 and 7.2 tasks on average, their median number of administered tasks ranged 5-6; students who started at task 3 were administered between 8.5 and 9.2 tasks on average, with a median number of administered tasks between 8 and 10. These data indicate that, for both these groups of students, the tendency was to administer more than the minimum number of tasks needed. Students who started at task 6 were

administered 6.9 to 7 tasks on average, with a median of 7 tasks. In other words, students who started at task 6 at were generally administered all 7 tasks available at the high-complexity level.

These results show that a large majority of the students assessed during the 2010 spring SC-Alt administration were administered at least the minimum number of tasks, and in many instances the test administrators exposed the students to additional, more complex, and more difficult tasks, beyond the minimal administration requirements.

Exhibit 8.8: Number of Tasks Administered, by Starting Task—ELA

Form = Elementary School (ES)													
		Number of Tasks Administered											
Starting Task		**<5**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**Total Students**	**Mean Number of Tasks**	**Median Number of Tasks**
Task 1	N	2	164	49	34	62	7	4	3	37	362		
	%	0.6	45.3	13.5	9.4	17.1	1.9	1.1	0.8	10.2	100	6.7	6
Task 3	N	0	0	8	63	11	21	221	—	—	324		
	%	0	0	2.5	19.4	3.4	6.5	68.2	—	—	100	9.2	10
Task 6	N	1	2	3	697	—	—	—	—	—	703		
	%	0.1	0.3	0.4	99.1	—	—	—	—	—	100	7.0	7
Total	N	3	166	60	794	73	28	225	3	37	1389		
	%	0.2	12	4.3	57.2	5.3	2	16.2	0.2	2.7	100	7.4	7
Form = Middle School (MS)													
		Number of Tasks Administered											
Starting Task		**<5**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**Total Students**	**Mean Number of Tasks**	**Median Number of Tasks**
Task 1	N	5	105	68	20	6	8	10	4	49	275		
	%	1.8	38.2	24.7	7.3	2.2	2.9	3.6	1.5	17.8	100	7.1	6
Task 3	N	0	0	1	45	10	14	157	—	—	227		
	%	0	0	0.4	19.8	4.4	6.2	69.2	—	—	100	9.2	10
Task 6	N	1	1	3	678	—	—	—	—	—	683		
	%	0.1	0.1	0.4	99.3	—	—	—	—	—	100	7.0	7
Total	N	6	106	72	743	16	22	167	4	49	1185		
	%	0.5	8.9	6.1	62.7	1.4	1.9	14.1	0.3	4.1	100	7.4	7
Form = High School (HS)													
		Number of Tasks Administered											
Starting Task		**<5**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**Total Students**	**Mean Number of Tasks**	**Median Number of Tasks**
Task 1	N	6	48	9	4	4	7	5	0	18	101		
	%	5.9	47.5	8.9	4	4	6.9	5	0	17.8	100	6.9	5
Task 3	N	1	0	0	9	5	5	32	—	—	52		
	%	1.9	0	0	17.3	9.6	9.6	61.5	—	—	100	9.1	10
Task 6	N	2	0	5	196	—	—	—	—	—	203		
	%	1	0	2.5	96.6	—	—	—	—	—	100	6.9	7
Total	N	9	48	14	209	9	12	37	0	18	356		
	%	2.5	13.5	3.9	58.7	2.5	3.4	10.4	0	5.1	100	7.3	7

Exhibit 8.9: Number of Tasks Administered, by Starting Task—Mathematics

Form = Elementary School (ES)													
		Number of Tasks Administered											
Starting Task		**<5**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**Total Students**	**Mean Number of Tasks**	**Median Number of Tasks**
Task 1	N	2	171	52	27	23	15	40	8	30	368		
	%	0.5	46.5	14.1	7.3	6.3	4.1	10.9	2.2	8.2	100	6.9	6
Task 3	N	1	0	2	74	136	22	122	—	—	357		
	%	0.3	0	0.6	20.7	38.1	6.2	34.2	—	—	100	8.5	8
Task 6	N	0	3	2	655	—	—	—	—	—	660		
	%	0	0.5	0.3	99.2	—	—	—	—	—	100	7.0	7
Total	N	3	174	56	756	159	37	162	8	30	1385		
	%	0.2	12.6	4	54.6	11.5	2.7	11.7	0.6	2.2	100	7.4	7
Form = Middle School (MS)													
		Number of Tasks Administered											
Starting Task		**<5**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**Total Students**	**Mean Number of Tasks**	**Median Number of Tasks**
Task 1	N	6	204	17	8	3	1	4	6	6	255		
	%	2.4	80	6.7	3.1	1.2	0.4	1.6	2.4	2.4	100	5.5	5
Task 3	N	2	1	2	68	39	33	120	—	—	265		
	%	0.8	0.4	0.8	25.7	14.7	12.5	45.3	—	—	100	8.7	9
Task 6	N	3	0	0	659	—	—	—	—	—	662		
	%	0.5	0	0	99.5	—	—	—	—	—	100	7.0	7
Total	N	11	205	19	735	42	34	124	6	6	1182		
	%	0.9	17.3	1.6	62.2	3.6	2.9	10.5	0.5	0.5	100	7.1	7
Form = High School (HS)													
		Number of Tasks Administered											
Starting Task		**<5**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**Total Students**	**Mean Number of Tasks**	**Median Number of Tasks**
Task 1	N	7	62	4	8	2	2	2	4	7	98		
	%	7.1	63.3	4.1	8.2	2	2	2	4.1	7.1	100	6.1	5
Task 3	N	0	0	1	15	3	7	36	—	—	62		
	%	0	0	1.6	24.2	4.8	11.3	58.1	—	—	100	9.0	10
Task 6	N	1	2	2	189	—	—	—	—	—	194		
	%	0.5	1	1	97.4	—	—	—	—	—	100	7.0	7
Total	N	8	64	7	212	5	9	38	4	7	354		
	%	2.3	18.1	2	59.9	1.4	2.5	10.7	1.1	2	100	7.1	7

Exhibit 8.10: Number of Tasks Administered, by Starting Task—Science

Form = Elementary School (ES)													
		Number of Tasks Administered											
Starting Task		**<5**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**Total Students**	**Mean Number of Tasks**	**Median Number of Tasks**
Task 1	N	3	132	87	4	20	22	5	2	25	300		
	%	1	44	29	1.3	6.7	7.3	1.7	0.7	8.3	100	6.5	6
Task 3	N	1	0	2	85	18	12	146	—	—	264		
	%	0.4	0	0.8	32.2	6.8	4.5	55.3	—	—	100	8.8	10
Task 6	N	1	0	4	378	—	—	—	—	—	383		
	%	0.3	0	1	98.7	—	—	—	—	—	100	7.0	7
Total	N	5	132	93	467	38	34	151	2	25	947		
	%	0.5	13.9	9.8	49.3	4	3.6	15.9	0.2	2.6	100	7.3	7
Form = Middle School (MS)													
		Number of Tasks Administered											
Starting Task		**<5**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**Total Students**	**Mean Number of Tasks**	**Median Number of Tasks**
Task 1	N	4	88	30	28	14	1	9	1	25	200		
	%	2	44	15	14	7	0.5	4.5	0.5	12.5	100	6.7	6
Task 3	N	0	0	2	23	32	3	109	—	—	169		
	%	0	0	1.2	13.6	18.9	1.8	64.5	—	—	100	9.1	10
Task 6	N	3	1	3	434	—	—	—	—	—	441		
	%	0.7	0.2	0.7	98.4	—	—	—	—	—	100	7.0	7
Total	N	7	89	35	485	46	4	118	1	25	810		
	%	0.9	11	4.3	59.9	5.7	0.5	14.6	0.1	3.1	100	7.4	7
Form = High School (HS)													
		Number of Tasks Administered											
Starting Task		**<5**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**Total Students**	**Mean Number of Tasks**	**Median Number of Tasks**
Task 1	N	5	63	13	5	5	3	0	3	14	111		
	%	4.5	56.8	11.7	4.5	4.5	2.7	0	2.7	12.6	100	6.4	5
Task 3	N	1	1	0	16	6	9	28	0	0	61		
	%	1.6	1.6	0	26.2	9.8	14.8	45.9	0	0	100	8.7	9
Task 6	N	1	0	7	175	—	—	—	—	—	183		
	%	0.5	0	3.8	95.6	—	—	—	—	—	100	6.9	7
Total	N	7	64	20	196	11	12	28	3	14	355		
	%	2	18	5.6	55.2	3.1	3.4	7.9	0.8	3.9	100	7.1	7

Exhibit 8.11: Number of Tasks Administered, by Starting Task—Social Studies

Form = Elementary School (ES)													
		Number of Tasks Administered											
Starting Task		**<5**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**Total Students**	**Mean Number of Tasks**	**Median Number of Tasks**
Task 1	N	0	107	38	12	5	25	6	3	43	239		
	%	0	44.8	15.9	5	2.1	10.5	2.5	1.3	18	100	7.2	6
Task 3	N	0	1	1	67	19	10	94	—	—	192		
	%	0	0.5	0.5	34.9	9.9	5.2	49	—	—	100	8.7	9
Task 6	N	5	2	4	508	—	—	—	—	—	519		
	%	1	0.4	0.8	97.9	—	—	—	—	—	100	6.9	7
Total	N	5	110	43	587	24	35	100	3	43	950		
	%	0.5	11.6	4.5	61.8	2.5	3.7	10.5	0.3	4.5	100	7.4	7

Form = Middle School (MS)													
		Number of Tasks Administered											
Starting Task		**<5**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**Total Students**	**Mean Number of Tasks**	**Median Number of Tasks**
Task 1	N	2	75	45	5	15	9	9	14	13	187		
	%	1.1	40.1	24.1	2.7	8	4.8	4.8	7.5	7	100	6.9	6
Task 3	N	0	0	2	27	30	24	44	—	—	127		
	%	0	0	1.6	21.3	23.6	18.9	34.6	—	—	100	8.6	9
Task 6	N	1	1	3	498	—	—	—	—	—	503		
	%	0.2	0.2	0.6	99	—	—	—	—	—	100	7.0	7
Total	N	3	76	50	530	45	33	53	14	13	817		
	%	0.4	9.3	6.1	64.9	5.5	4	6.5	1.7	1.6	100	7.2	7

Number of Items Administered

Since test administrators were instructed to administer all of the items in a task and each task contained approximately five items, the number of items administered was roughly proportional to the number of tasks administered. Exhibits 8.12 through 8.15 show the mean, median, and 25th and 75th percentile for number of administered items, disaggregated by content area, form level, and task start-point.

The median number of items administered to students starting at task 1 ranged 23-33 across content areas and form levels, the median for task 3 start-points ranged 38-55, and the median for task 6 start-points ranged 30-40. Students beginning at task 6 were administered fewer and a smaller range of items than students starting at Task 3 since these students demonstrated more predictable performance (according to the SPQ results) and the end of the minimally required task range coincided with the end of the test. Students whose test administration began at task 3 tended to have more items administered to them. These were the students whose performance levels were the least predictable, thus administering more items and testing these students more extensively provided them with optimal opportunities to demonstrate their proficiency. In these cases, the administration of a content area test was ended only when (1) the end of the test was reached or (2) when the student could no longer respond successfully on a task (i.e., failed to obtain three or more points on the task).

Exhibit 8.12: Number of Items Administered, by Starting Task—ELA

Form = Elementary School					
Starting Task	**N**	**Mean**	**P25**	**Median**	**P75**
Task 1	362	36.9	28	33	43
Task 3	324	50.4	44	55	55
Task 6	703	36.9	37	37	37
Total	1389	40.1	37	37	43
Form = Middle School					
Starting Task	**N**	**Mean**	**P25**	**Median**	**P75**
Task 1	275	38.5	29	33	49
Task 3	227	46.6	45	50	50
Task 6	683	32.9	33	33	33
Total	1185	36.9	33	33	37
Form = High School					
Starting Task	**N**	**Mean**	**P25**	**Median**	**P75**
Task 1	101	36.5	26	26	47
Task 3	52	47.8	42	53	53
Task 6	203	37.7	38	38	38
Total	356	38.8	38	38	38

Exhibit 8.13: Number of Items Administered, by Starting Task—Mathematics

Form = Elementary School					
Starting Task	**N**	**Mean**	**P25**	**Median**	**P75**
Task 1	368	35.6	28	32	44
Task 3	357	40.7	38	38	48
Task 6	660	29.9	30	30	30
Total	1385	34.2	30	30	38
Form = Middle School					
Starting Task	**N**	**Mean**	**P25**	**Median**	**P75**
Task 1	255	25.7	23	23	23
Task 3	265	42.7	35	43	49
Task 6	662	35.9	36	36	36
Total	1182	35.2	35	36	36
Form = High School					
Starting Task	**N**	**Mean**	**P25**	**Median**	**P75**
Task 1	98	29.9	24	24	34
Task 3	62	49.2	38	55	55
Task 6	194	39.7	40	40	40
Total	354	38.7	38	40	40

Exhibit 8.14: Number of Items Administered, by Starting Task—Science

Form = Elementary School					
Starting Task	**N**	**Mean**	**P25**	**Median**	**P75**
Task 1	300	30.8	23	28	34
Task 3	264	43.4	34	50	50
Task 6	383	36.9	37	37	37
Total	947	36.8	34	37	38
Form = Middle School					
Starting Task	**N**	**Mean**	**P25**	**Median**	**P75**
Task 1	200	37.1	29	33	41
Task 3	169	47.2	42	51	51
Task 6	441	32.8	33	33	33
Total	810	36.9	33	33	39
Form = High School					
Starting Task	**N**	**Mean**	**P25**	**Median**	**P75**
Task 1	111	32	26	26	35
Task 3	61	39.7	33	41	45
Task 6	183	29.8	30	30	30
Total	355	32.2	30	30	33

Exhibit 8.15: Number of Items Administered, by Starting Task—Social Studies

Form = Elementary School					
Starting Task	**N**	**Mean**	**P25**	**Median**	**P75**
Task 1	239	31.6	20	24	41
Task 3	192	42.3	33	44	50
Task 6	519	37.7	38	38	38
Total	950	37.1	37	38	38
Form = Middle School					
Starting Task	**N**	**Mean**	**P25**	**Median**	**P75**
Task 1	187	30.5	22	26	36
Task 3	127	39.7	37	41	46
Task 6	503	31.9	32	32	32
Total	817	32.8	32	32	32

Achievement Level of Students by Start Point

Within an SC-Alt form, two or more tasks (consisting of an average of five items each) are used to assess the same standards at different levels of student communication and content complexity and are placed on the test form in locations that ensure that there is adequate content coverage of the standards regardless of the student's starting point in the assessment. Although tasks are ordered on the form based on student communication levels and average content complexity, items of both lower and higher complexity may appear in each task. This configuration presents items and tasks across the entire assessment providing students with opportunities to demonstrate proficiency. Each student's proficiency and resulting achievement level are determined by the student's performance on the specific group of items the student was administered. The calculation of student proficiency scores is described in Chapter 5. The distribution of

achievement levels for students according to task start point, form level, and content area is presented in Exhibit 8.16.

The table entries demonstrate interesting operational aspects of the leveled structure of the SC-Alt. Across content areas, students beginning the assessment at Task 1 are categorized as Proficient (achievement levels 3 and 4) at rates between 3% and 45%, with the lowest percentage in middle school (3%–13%), most varied in elementary school (11%–45%), and most consistent in high school (8%–15%). For students starting at Task 3, 26% to 95% of students across content areas are categorized as Proficient; as with students beginning with Task 1, the lowest percentage of Proficient students starting at Task 3 is demonstrated in middle school (26%–65%), a wide range is shown in elementary school (45%–95%), while high school (43%–75%) is most consistent. Finally, 84% to 99% of students starting at Task 6 are Proficient with the range of percentages becoming narrower across test forms, unlike the other task start-points.

Summary

The purpose of the start and stop point analyses was to document the number of tasks and items students complete during the assessment and the performance levels that groups of students attain who enter the assessment at different start-points. The results of these analyses demonstrate the effectiveness of the use of the SPQ and the test administration start and stop rules that are based on the student's performance during the assessment. Except in a few instances, all students were administered adequate numbers of tasks and items to assess the intended content.

The findings indicate that SPQ start/stop rules were being followed for almost all tested students. A considerable proportion of tested students continued testing beyond the minimum required number of tasks to be administered. As a consequence, in each starting task group, there were students who tested in the proficient range (i.e., at achievement levels 3 or 4). Finally, students assigned to higher starting tasks showed a greater likelihood of testing in the proficient range. These results demonstrate that the tailored assessment design of the SC-Alt operates as expected.

Exhibit 8.16: Achievement Level by Task Start Point, Form Level, and Content Area

		Elementary School (ES)								Middle School (MS)								High School (HS)							
		Starting Task								Starting Task								Starting Task							
		1		3		6		Total		1		3		6		Total		1		3		6		Total	
Subject	Achievement Level	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
ELA	Level 1	75	20.7	1	0.3	0	0.0	76	5.5	74	26.9	3	1.3	0	0.0	77	6.5	41	40.6	0	0.0	0	0.0	41	11.5
	Level 2	185	51.1	60	18.5	7	1.0	252	18.1	165	60.0	77	33.9	25	3.7	267	22.5	45	44.6	13	25.0	8	3.9	66	18.5
	Level 3	83	22.9	147	45.4	81	11.5	311	22.4	32	11.6	80	35.2	82	12.0	194	16.4	11	10.9	15	28.8	39	19.2	65	18.3
	Level 4	19	5.2	116	35.8	615	87.5	750	54.0	4	1.5	67	29.5	575	84.2	646	54.5	4	4.0	24	46.2	156	76.8	184	51.7
	Proficient	102	28.2	263	81.2	696	99.0	1061	76.4	36	13.1	147	64.8	657	96.2	840	70.9	15	14.9	39	75.0	195	96.1	249	69.9
Math	Level 1	100	27.2	1	0.3	0	0.0	101	7.3	118	46.3	4	1.5	0	0.0	122	10.3	56	57.1	1	1.6	0	0.0	57	16.1
	Level 2	205	55.7	120	33.6	7	1.1	332	24.0	129	50.6	144	54.3	26	3.9	299	25.3	34	34.7	31	50.0	24	12.4	89	25.1
	Level 3	61	16.6	210	58.8	217	32.9	488	35.2	7	2.7	104	39.2	256	38.7	367	31.0	6	6.1	25	40.3	106	54.6	137	38.7
	Level 4	2	0.5	25	7.0	436	66.1	463	33.4	1	0.4	12	4.5	379	57.3	392	33.2	2	2.0	5	8.1	64	33.0	71	20.1
	Proficient	63	17.1	235	65.8	653	98.9	951	68.7	8	3.1	116	43.8	635	95.9	759	64.2	8	8.2	30	48.4	170	87.6	208	58.8
Science	Level 1	65	21.7	1	0.4	0	0.0	66	7.0	108	54.0	5	3.0	0	0.0	113	14.0	76	68.5	8	13.1	0	0.0	84	23.7
	Level 2	99	33.0	12	4.5	1	0.3	112	11.8	66	33.0	61	36.1	22	5.0	149	18.4	23	20.7	27	44.3	29	15.8	79	22.3
	Level 3	95	31.7	85	32.2	18	4.7	198	20.9	20	10.0	55	32.5	68	15.4	143	17.7	5	4.5	19	31.1	52	28.4	76	21.4
	Level 4	41	13.7	166	62.9	364	95.0	571	60.3	6	3.0	48	28.4	350	79.4	404	49.9	7	6.3	7	11.5	102	55.7	116	32.7
	Proficient	136	45.3	251	95.1	382	99.7	769	81.2	26	13.0	103	60.9	418	94.8	547	67.5	12	10.8	26	42.6	154	84.2	192	54.1
Social Studies	Level 1	105	43.9	6	3.1	1	0.2	112	11.8	90	48.1	5	3.9	0	0.0	95	11.6								
	Level 2	109	45.6	100	52.1	34	6.6	243	25.6	90	48.1	89	70.1	59	11.7	238	29.1								
	Level 3	20	8.4	75	39.1	254	48.9	349	36.7	6	3.2	29	22.8	198	39.4	233	28.5								
	Level 4	5	2.1	11	5.7	229	44.1	245	25.8	1	0.5	4	3.1	244	48.5	249	30.5								
	Proficient	25	10.5	86	44.8	483	93.1	594	62.5	7	3.7	33	26.0	442	87.9	482	59.0								

Chapter 9: High School Biology Field Test

A high-school biology assessment was field-tested in spring 2010. The primary purpose of this field test was to pilot a pool of tasks and items, compute item statistics, and select items to appear on the 2011 spring operational assessment.

This new Biology assessment is based on new content standards and indicators that are not articulated with the currently administered physical science assessments in elementary and middle school. Therefore, there is no attempt to establish a vertical scale linking the high school biology assessment to physical science.

The following sections detail information on the form design of the high school biology assessment, specifically describing the linking across field-test forms and the start point determination. The initial sampling plan is described, and characteristics of the realized, actual, sample are compared with the intended sampling frame. Subsequent to administration, item statistics were calculated, and the results of an Item Data Review meeting are reported. Finally, a meeting was convened to set performance standards in high school biology using the ID Matching procedure (Ferrara, Perie, & Johnson, 2008). A summary of the meeting and resulting recommendations are presented here; full details of the standard-setting meeting can be found in American Institutes for Research and South Carolina Department of Education (2010).

Forms and Linking Design

This section describes the within-level, cross-form linking design and test administration procedures for the SC-Alt High School Biology Field Test in spring 2010. The scale of the biology assessment is not intended to be linked to other grade-bands.

The High School Biology Field Test included 21 tasks in two field-test booklets—each containing nine unique tasks and three linking tasks. Tasks were arranged in order of target complexity. Along with the test booklet, test administrators received

- a Test Administrator's Manual (TAM);
- a manipulatives kit (with printed and physical manipulatives for all tasks);
- a student answer folder for all participating students; and
- a draft Student Placement Questionnaire (SPQ).

Forms Design of the Biology Field Test

The field-test forms were constructed to be similar in structure and organization to the current operational forms in each content area with a total of 12 tasks. Twenty-five tasks, and approximately 175 items, were submitted to SCDE and its content and fairness committee for review. Of those, 21 tasks and associated items were selected for field-testing.

As illustrated in Exhibit 9.1, the Biology Field Test includes two field-test forms—each containing nine unique tasks and three linking tasks. The tasks for each of the two field-test forms were arranged in increasingly complex order as judged by SCDE, its content committee, and AIR content staff.

Exhibit 9.1: Numbers of Tasks in Each Field-Test Form

	Tasks in Each Test Form		
Form	**Unique Tasks**	**Linking Tasks**	**Total Tasks Per Form**
Form A	9	3	12
Form B	9		12
Total	18	3	21

Administration rules and guidelines for the field-test forms paralleled those used for the operational assessment. A similar test structure and procedures consistent with the operational assessments were maintained throughout the field test.

Linking Task and Item Development Approach

In order to improve the likelihood of stable linking constants between forms, the linking tasks and items were designed with two major characteristics in mind: content spread and difficulty spread. The strategy was to define the link by a number of well-fitting and stable linking items.

To the maximum degree possible, the items in the linking tasks reflected the content standards being assessed. Individually and collectively, the linking tasks contained items that best represented the academic content standards for high school biology. Admittedly this spread was a bit of a challenge to achieve, given that each task typically contained about seven items, and there were six academic standards and 24 extended indicators eligible for inclusion on the assessment. While item development was focused at the academic standard level, every effort was made to include as many of the 24 indicators in the approximately 21 linking items as possible.

Similarly, the distribution of the difficulty of the linking items was generally aimed toward the distribution of the content trait in the student sample. Linking items were developed so as not to be too difficult or too easy for the sampled students. Rather, as a priori content analyses permitted, the linking tasks were accessible (from a difficulty and task complexity perspective) by virtually all students in the sample.

Linking Task Positions

The three linking tasks were included in the same positions on each of the two test forms, at positions three, six, and nine in the test booklets. This placement maximizes the number of responses per linking task, given the starting tasks of 1 and 3, and the guidelines for the minimum number of tasks to be administered. The linking tasks were positioned on the test forms according to expected difficulty; the linking task judged to be the least difficult was placed in position three, the linking task of intermediate difficulty was placed in position six, and the most difficult task was placed in position nine.

The linking tasks and associated items were also targeted to general levels of cognitive functioning. The task at position three is primarily focused toward students at the pre-symbolic and concrete levels. The task at position six is focused at the concrete level, and the task at position nine is focused at the concrete and abstract levels.

Starting Point for the Biology Field Test

There were two starting points for the High School Biology Field Test. Test administrators were instructed to begin the assessment at either Task 1 or Task 3 according to the following guidelines:

- For 15-year-old students, the starting point for the Biology Field Test was determined according to their 2010 physical science assessment.
  - If a student started at Task 1 or 3 of the physical science assessment, he or she started at Task 1 on the Biology Field Test. The teacher then administered Tasks 1–9 to the student.
  - On the other hand, if the student started at Task 6 of the physical science assessment, then the student was started at Task 3 on the Biology Field Test. In this case, the teacher was instructed to administer Tasks 3–12 to the student.
- For 16-year-old students, their starting points were determined based on their 2009 physical science assessment.
  - Like the 15-year-old students, if the student started at Task 1 or 3 on the 2009 physical science assessment, then the student was to start at Task 1 and be administered Tasks 1–9.
  - If the 16-year-old student started at Task 6 on the physical science assessment, then the student was to start at Task 3 on the Biology Field Test and be administered Tasks 3–12.

The student's starting point was indicated on the student roster delivered with the testing materials.

Sampling

Two forms of the High School Biology Field Test were administered during the spring 2010 administration. In order to obtain sample sizes sufficient to ensure at least 250 responses per item, all teachers of grade 10 students participated in the field test.

The teachers were instructed to administer the High School Biology Field Test to

- all grade 10 students 15 years old by September 1, 2009, and
- all students 16 years old who took the high school (grade 10) science assessment in spring 2009 as long as their teacher/test administrator was also tasked to administer the SC-Alt to other students in 2010 spring (i.e., operational administration).

In combination, the two groups of students should provide a sufficient sample to result in the target number of responses per item.

The following guidelines were employed to assign Biology Field Test forms A and B:

- The Biology Field Test forms were assigned by school. Each school was sent only one form (either Form A or Form B) for the assessment. The teachers at the school were

responsible for administering that form. Materials were boxed by school and sent to each of the school districts in South Carolina.

- Test forms for the 15-year-old students were assigned based on the field-test precode data. Test forms were assigned according to a stratified random sampling method so that each form would be (1) taken by approximately 50% of the population and (2) distributed about equally among students with regard to primary disabilities.
- Sixteen-year-old students participating in the High School Biology Field Test were to take the specific form that had been assigned to their school as described above.

Based on the 2010 spring precode data, information regarding the intended Biology Field Test sample is provided below. Following these guidelines, the distribution of students, teachers, schools, and districts by field-test form (A or B) is presented in Exhibit 9.2. The numbers of students per school assigned to Biology Field Test forms A and B are described in Exhibit 9.3. Finally, demographics for the intended sample are described by form in Exhibit 9.4.

Exhibit 9.2: Distribution of Students, Teachers, Schools, and Districts by Form

Number	Form A	Form B	Total
Students	221	268	489
Teachers	112	126	238
Schools	92	95	187
Districts	53	50	103

Exhibit 9.3: Distribution of Students per School by Form

	Form A		Form B		Total	
Students per School	**N**	**%**	**N**	**%**	**N**	**%**
1	44	47.83	44	46.32	88	47.06
2	16	17.39	13	13.68	29	15.51
3	14	15.22	14	14.74	28	14.97
4	6	6.52	7	7.37	13	6.95
5	5	5.43	5	5.26	10	5.35
6	2	2.17	6	6.32	8	4.28
7	2	2.17	2	2.11	4	2.14
8	1	1.09	1	1.05	2	1.07
9	1	1.09	.	.	1	0.53
11	1	1.09	.	.	1	0.53
12	.	.	1	1.05	1	0.53
15	.	.	1	1.05	1	0.53
18	.	.	1	1.05	1	0.53

Exhibit 9.4: Summary of Demographic Information

	Form A		Form B		Total	
	N	%	N	%	N	%
STUDENT'S ETHNICITY						
African American	105	47.5	136	50.8	241	49.3
African American/American Indian	.	0.0	1	0.4	1	0.2
American Indian	1	0.5	.	0.0	1	0.2
Asian	4	1.8	2	0.8	6	1.2
Hispanic	7	3.2	4	1.5	11	2.2
White	102	46.2	122	45.5	224	45.8
White/African American	2	0.9	2	0.8	4	0.8
Other	.	0.0	1	0.4	1	0.2
STUDENT'S GENDER						
Female	78	35.3	96	35.8	174	35.6
Male	143	64.7	172	64.2	315	64.4
ESL (LANGUAGE)						
English Speaker II	215	97.3	263	98.1	478	97.8
Pre-functional	6	2.7	4	1.5	10	2.0
Unknown	.	0.0	1	0.4	1	0.2
EFA GRADE (REPORTED GRADE FOR FUNDING)						
5	5	2.3	3	1.1	8	1.6
6	1	0.5	3	1.1	4	0.8
7	6	2.7	4	1.5	10	2.0
8	41	18.6	11	4.1	52	10.6
9	56	25.3	77	28.7	133	27.2
10	85	38.5	129	48.1	214	43.8
11	23	10.4	35	13.1	58	11.9
12	4	1.8	6	2.2	10	2.0
STUDENT AGE						
15	159	72.0	169	63.1	328	67.1
16	62	28.1	99	36.9	161	32.9
IEP DISABILITY CODES (MULTIPLE CODES PER STUDENT)						
Severely Mentally Disabled	22	10.0	29	10.8	51	10.4
Moderately Mentally Disabled	103	46.6	112	41.8	215	44.0
Mildly Mentally Disabled	36	16.3	57	21.3	93	19.0
Autism	40	18.1	45	16.8	85	17.4
Deaf/Blindness	.	0.0	7	2.6	7	1.4
Emotional Disability	.	0.0	1	0.4	1	0.2

	Form A		Form B		Total	
	N	%	N	%	N	%
Hearing Impaired	3	1.4	8	3.0	11	2.2
Learning Disability	.	0.0	3	1.1	3	0.6
Multiple-Disability	8	3.6	21	7.8	29	5.9
Other Health Impaired	12	5.4	11	4.1	23	4.7
Orthopedically Impaired	8	3.6	12	4.5	20	4.1
Speech or Language Impaired	70	31.7	92	34.3	162	33.1
Traumatic Brain Injury	.	0.0	2	0.8	2	0.4
Visually Impaired	3	1.4	15	5.6	18	3.7
TOTAL	221	100.0	268	100.0	489	100.0

Analysis, Calibration and Linking

All analyses for the 2010 spring High School Biology Field Test parallel those previously conducted for the 2006 fall Science Field Test. The analyses were conducted in four steps:

1. Data preparation and quality control
2. Classical item analyses (including DIF statistics)
3. Data review for items flagged for statistical or DIF criteria
4. Item Response Theory (IRT) calibration for use in operational form construction

Item Data Review

Items flagged on the basis of criteria described in American Institutes for Research and South Carolina Department of Education (2010) were reviewed by AIR psychometricians and SCDE officials. First, a team of AIR psychometricians reviewed all flagged items to ensure that the data were accurate and properly analyzed, that response keys were correct, and that there were no obvious problems with the items. AIR recommended whether the item should be retained in the item pool or discarded, depending on the reason for the item flag and its effect on the quality of the assessment as a whole. SCDE had the final authority on whether the flagged items should be included in the operational scoring, based on the item statistics and content appropriateness of the items.

Following review of flagged items, eight biology items were rejected and, therefore, excluded from operational form construction:

- Two items were rejected on the basis of low polyserial correlations (Item ID 1444 and 1475).
- Three items were rejected based on poor performance of multiple statistics (Item ID 1548, 1451, and 1468).
- Three items were rejected based on detailed review of item statistics and content (Item ID 1446, 1456, and 1457).

Item Response Theory (IRT) Calibration and Creation of a Common Difficulty Scale (Linking)

Field-test data were calibrated according to Item Response Theory (IRT) according to Masters' Partial Credit model (Masters, 1982) using the Winsteps software. A common task/item design, as described in Exhibit 9.1, was used, allowing for common item equating of the two biology field-test forms. Using the common item design, field-test items are jointly calibrated in a single Winsteps run, placing the item parameters for both field-test forms on the same scale. AIR's internal analyses showed substantially less linking error when the forms were calibrated jointly than when other linking methods were used.

Item statistics for the spring 2010 High School Biology Field Test forms can be found in Appendix F.

Demographics of Participating Students

The intended sample described under "Sampling" is compared with the actual administration results in Exhibits 9.5, 9.6, and 9.7.

- The actual administration achieved 97% (n = 472) of the intended total number of students (n = 489) and closely approximated the target of 250 students per test form (Form A: n = 235; Form B = 237).
- The actual administration included fewer teachers (82% of intended sample), schools (78%), and districts (83%) than included in the intended sample.
- Forms A and B were nearly equally distributed across the students, teachers, schools, and districts in the actual administration.
- The majority of students in the intended sample (78%) and actual administration (68%) were assessed in schools with three students or fewer participating in the Biology Field Test. The distribution of assessed students when the school contained more than three Biology Field Test students was similar for the intended and actual administration.
- Demographics of the actual administration very closely approximated the intended sample with respect to ethnicity, gender, English language learners, grade, and student age. Free and reduced lunch information was not included in the Biology Field Test sampling frame.
- Proportions of student in the actual administration classified as severely (11%), moderately (44%), or mildly mentally disabled (19%) were similar to the proportions specified in the intended sample (10%, 44%, and 19% respectively).
- Autism classification representation in the assessed sample (14.6) was also similar to the proportion in the intended sample (17.4)
- Other disability classifications were similarly distributed in actual administration when compared with the intended sample; of these disability codes, speech or language impairment was the largest disability category in both the actual (31%) and intended samples (32%), while all other disability categories were assigned to less than 5% of students in either sample.

Exhibit 9.5: Distribution of Students, Teachers, Schools, and Districts by Form

	Intended Sample					Actual				
	Form A		Form B			Form A		Form B		
Number	N	%	N	%	TOTAL	N	%	N	%	TOTAL
Students	221	45	268	55	489	235	50	237	50	472
Teachers	112	47	126	53	238	95	49	99	51	194
Schools	92	49	95	51	187	79	54	66	46	145
Districts	53	51	50	49	103	49	57	37	43	86

Exhibit 9.6: Distribution of Students per School by Form

	Intended Sample					Actual Sample				
	Form A		Form B			Form A		Form B		
Students per School	N	%	N	%	TOTAL	N	%	N	%	TOTAL
1	44	48	44	46	88	25	32	23	34	48
2	16	17	13	14	29	13	16	11	16	24
3	14	15	14	15	28	17	22	10	15	27
4	6	6.5	7	7.4	13	8	10	4	6	12
5	5	5.4	5	5.3	10	6	7.6	6	9	12
6	2	2.2	6	6.3	8	6	7.6	3	4.5	9
7	2	2.2	2	2.1	4	2	2.5	3	4.5	5
8	1	1.1	1	1.1	2	.	.	2	3	2
9	1	1.1	.	.	1	1	1.3	1	1.5	2
11	1	1.1	.	.	1	.	.	1	1.5	1
12	.	.	1	1.1	1	1	1.3	1	1.5	2
15	.	.	1	1.1	1	.	.	1	1.5	1
17					0	.	.	1	1.5	1
18	.	.	1	1.1	1	.	.	.	.	0

Exhibit 9.7: Summary of Demographic Information

	Sample						Actual					
	A		B		TOTAL		A		B		TOTAL	
	N	%	N	%	N	%	N	%	N	%	N	%
STUDENT'S ETHNICITY												
African American	105	47.5	136	50.8	241	49.3	120	51.1	122	51.5	242	51.3
African American/American Indian		0	1	0.4	1	0.2	1	0.4		0	1	0.2
American Indian	1	0.5		0	1	0.2		0	1	0.4	1	0.2
Asian	4	1.8	2	0.8	6	1.2	2	0.9	2	0.8	4	0.8
Hawaiian/Pacific Islander		0		0	0	0.0		0		0	0	0.0
Hispanic	7	3.2	4	1.5	11	2.2	6	2.6	4	1.7	10	2.1

	Sample						Actual					
	A		B		TOTAL		A		B		TOTAL	
	N	%	N	%	N	%	N	%	N	%	N	%
White	102	46.2	122	45.5	224	45.8	101	43	107	45.2	208	44.1
White/African American	2	0.9	2	0.8	4	0.8	3	1.3	1	0.4	4	0.8
White/American Indian		0		0	0	0.0		0		0	0	0.0
White/Asian		0		0	0	0.0		0		0	0	0.0
Other		0	1	0.4	1	0.2	1	0.4		0	1	0.2
Unknown		0		0	0	0.0	1	0.4		0	1	0.2
STUDENT'S GENDER												
Female	78	35.3	96	35.8	174	35.6	74	31.5	85	35.9	159	33.7
Male	143	64.7	172	64.2	315	64.4	161	68.5	152	64.1	313	66.3
Blank		0		0	0	0.0		0		0	0	0.0
ESL (LANGUAGE)												
Advanced		0		0		0.0		0		0	0	0.0
Advanced Waiver		0		0		0.0		0		0	0	0.0
Beginner		0		0		0.0		0		0	0	0.0
Beginner Waiver		0		0		0.0		0		0	0	0.0
English Speaker I		0		0		0.0	1	0.4		0	1	0.2
English Speaker II	215	97.3	263	98.1	478	97.8	209	88.9	229	96.6	438	92.8
Full English Proficient		0		0		0.0		0		0	0	0.0
Intermediate		0		0		0.0		0		0	0	0.0
Intermediate Waiver		0		0		0.0		0		0	0	0.0
Pre-functional	6	2.7	4	1.5	10	2.0	6	2.6	3	1.3	9	1.9
Pre-functional Waiver		0		0		0.0		0		0	0	0.0
Title III First Year Exited		0		0		0.0		0		0	0	0.0
Title III Second+ Year Exited		0		0		0.0		0		0	0	0.0
Unknown		0	1	0.4	1	0.2	19	8.1	5	2.1	24	5.1
ELIGIBLE FOR FREE OR REDUCED-PRICE LUNCH												
Free Meals		0		0		0.0	148	63	141	59.5	289	61.2
Full-Pay Meals		0		0		0.0	64	27.2	81	34.2	145	30.7
Reduced Meals		0		0		0.0	22	9.4	15	6.3	37	7.8
Unknown	221	100	268	100	489	100.0	1	0.4		0	1	0.2
EFA GRADE(REPORTED GRADE FOR FUNDING)												
5	5	2.3	3	1.1	8	1.6	1	0.4	7	3	8	1.7
6	1	0.5	3	1.1	4	0.8		0	4	1.7	4	0.8
7	6	2.7	4	1.5	10	2.0	6	2.6	3	1.3	9	1.9
8	41	18.6	11	4.1	52	10.6	40	17	12	5.1	52	11.0
9	56	25.3	77	28.7	133	27.2	72	30.6	60	25.3	132	28.0

	Sample						Actual					
	A		B		TOTAL		A		B		TOTAL	
	N	%	N	%	N	%	N	%	N	%	N	%
10	85	38.5	129	48.1	214	43.8	85	36.2	116	49	201	42.6
11	23	10.4	35	13.1	58	11.9	24	10.2	31	13.1	55	11.7
12	4	1.8	6	2.2	10	2.0	6	2.6	4	1.7	10	2.1
Other		0		0		0.0	1	0.4		0	1	0.2
STUDENT AGE												
14		0		0		0.0	2	0.9	1	0.4	3	0.6
15	159	72	169	63.1	328	67.1	147	62.6	171	72.2	318	67.4
16	62	28.1	99	36.9	161	32.9	82	34.9	62	26.2	144	30.5
Other		0		0		0.0	4	1.7	3	1.3	7	1.5
IEP DISABILITY CODES (MULTIPLE CODES PER STUDENT)												
Severely Mentally Disabled	22	10	29	10.8	51	10.4	14	6	37	15.6	51	10.8
Moderately Mentally Disabled	103	46.6	112	41.8	215	44.0	106	45.1	102	43	208	44.1
Mildly Mentally Disabled	36	16.3	57	21.3	93	19.0	42	17.9	48	20.3	90	19.1
Autism	40	18.1	45	16.8	85	17.4	46	19.6	23	9.7	69	14.6
Deaf/Blindness		0	7	2.6	7	1.4	2	0.9	5	2.1	7	1.5
Emotional Disability		0	1	0.4	1	0.2	2	0.9		0	2	0.4
Hearing Impaired	3	1.4	8	3	11	2.2	3	1.3	7	3	10	2.1
Learning Disability		0	3	1.1	3	0.6	1	0.4		0	1	0.2
Multiple-Disability	8	3.6	21	7.8	29	5.9	5	2.1	23	9.7	28	5.9
Other Health Impaired	12	5.4	11	4.1	23	4.7	12	5.1	9	3.8	21	4.4
Orthopedically Impaired	8	3.6	12	4.5	20	4.1	6	2.6	14	5.9	20	4.2
Speech or Language Impaired	70	31.7	92	34.3	162	33.1	75	31.9	73	30.8	148	31.4
Traumatic Brain Injury		0	2	0.8	2	0.4	2	0.9		0	2	0.4
Visually Impaired	3	1.4	15	5.6	18	3.7	8	3.4	11	4.6	19	4.0
TOTAL	**221**	**100**	**268**	**100**	**489**	**100**	**235**	**100**	**237**	**100**	**472**	**100**

Demographics for the actual administration of the Biology Field Test are compared with the demographics for the high school operational assessment (Chapter 3; Exhibit 3.3). This comparison indicates that the results for the Biology Field Test are similar to the operational assessments in ELA, mathematics, and social studies.

- The ethnicity of the majority of students in the Biology Field Test were identified as either African American (51%) or white (44%), similar to the operational assessments (53% and 42%).
- Gender distribution in the Biology Field Test (female = 34%; male = 66%) was almost exactly the same as in the operational assessments.

- Most students were classified as "English Speaker II" across all assessments (biology field test = 93%; operational = 97%). A higher proportion of "Unknown" language classifications were present in biology (5%) than in the operational assessments (<1%).
- Most students who were administered the SC-Alt received free lunch (biology = 61%; operational = 59%), then full-pay meals (31% and 32%), and reduced-price meals (8% and 9%).
- The actual administration of the Biology Field Test achieved smaller proportions of students in the EFA grades of 5 through 9 (funding grades, usually determined by program location) than the operational assessments. Conversely, greater proportions of students in grades 10 (43%), 11 (12%), and 12 (2%) were assessed by the Biology Field Test than by the operational assessments (39%, 3%, and 1%).
- Distributions of classifications of severely, moderately, and mildly mentally disabled were similar across the Biology Field Test and the operational assessments. The remaining disability classifications show proportions in the field test similar to the operational assessments, with minor fluctuations; "speech or language impaired" is the most common classification in both the field test (31%) and operational assessments (35%) and "multiple-disability" (6% and 5%) and "other health impaired" (4% and 5%) are the next most common in classifications in both assessments.

Standard Setting

On September 14 and 15, the American Institutes for Research (AIR), under contract to the South Carolina Department of Education (SCDE), convened a diverse panel of 19 educators and non-educators to recommend status performance standards on the 2010 spring field-test administration data for high school biology.

Using the Item Descriptor (ID) Matching method (see Cizek & Bunch, 2007; Ferrara, Perie, & Johnson, 2008), the panelists reviewed test items and the corresponding Description of Achievement Levels (DALs). They then recommended achievement standards for four levels (Levels 1, 2, 3, and 4), where Level 3 and above indicates proficiency. These standards were translated into cut points on the student proficiency scale by AIR psychometricians.

A summary of the standard-setting meeting and results follows; a detailed report can be found in American Institutes for Research and South Carolina Department of Education (2010).

The ID Matching Standard-Setting Process

The ID Matching standard-setting process, as described in the Standard-Setting Plan submitted to SCDE and reviewed by the South Carolina Technical Advisory Committee, was used at a workshop in Columbia on September 14 and 15, 2010, with a panel of 19 members. AIR staff provided training and led the participants through two rounds of ID Matching to first set the Level 3 (Proficient) performance standard and then the Level 2 and Level 4 performance standards.

Before the participants made each of their recommendations using the ID Matching procedure, they were given a readiness form to ensure that they fully understood the task and were prepared to place the performance standard. Analysis of these evaluations showed unanimous agreement

from the participants that they understood the task and were prepared to make performance standard recommendations.

Goals of the Standard Setting

The goals of the meeting, as stated to the panelists, were to

- recommend performance standards for High School Biology that correspond to the DALs for Levels 1 through 4;
- consider the agreement and impact data to guide judgments about item difficulty and placement of the performance standards; and
- recommend to SCDE the appropriate placement of cut points on the student proficiency scales for the High School Biology portion of SC-Alt.

Panel Composition

The recruiting plan for obtaining panelists for the standard setting workshops was intended to result in representative groups of panelists that would render informed recommendations to the South Carolina Department of Education on the placement of the cut scores for High School Biology. A diverse group of panelists brings a wide range of perspectives and experience to the standard setting effort, ensuring that the recommendations that are forwarded to the SCDE will be thoughtful and representative of broad educational constituencies.

SCDE recruited the panelists for the standard setting workshop. 15 panelists were targeted to participate in the panel and 19 panelists were recruited and attended the standard setting meeting. A demographic breakdown of target recruitment and the actual recruitment of the 19 panelists is presented in Table 4. Prior to setting any performance standards, 2 of the panelists withdrew leaving a total of 17 panelists to determine the cut scores.

Exhibit 9.8: Composition of the Standard Setting Panel

CHARACTERISTIC	TARGET	ACTUAL
TEACHERS	**8 Total**	**9 Total**
Spec Edc Teachers of High School Age Students– (representative of disability groups participating in alternate assessment)	6	7
General Education Biology/Science Teachers	2	2
OTHER EDUCATORS	**5 Total**	**8 Total**
Special Education Administrator	1	5
District Test Coordinator for SC-Alt	1	1
ESOL Specialists	2	2
Biology/Science Curriculum Specialists	1	1
COMMUNITY	**2 Total**	**2 Total**
Parents (parent of student with disabilities)	1	1
Department of Disabilities and Special Needs	1	
Autism Society	1	
Pro Parents of South Carolina	1	1
GENDER		
Female	8	14
Male	7	5
ETHNICITY		
African American	7	3
White	7	16
Other (Hispanic, Asian, or Native American)	1	
REGION (Small Rural and Metropolitan Areas)		
Midlands	9	12
Up-Country	3	5
Low-Country	3	2
TOTAL	**15**	**19**

The panelists were a diverse group; all participants had a background in education, with experience as classroom teachers, special education teachers, administrators. Several panelists had varied experience with special education and students with disabilities and as parents, and thus were able to represent a broad spectrum of perspectives. Three panelists also served on the State Alternate Assessment Advisory Committee.

Mechanics of the Standard-Setting Workshop

Workshop participants selected performance standards for the High School Biology test during two rounds of consideration with the DALs. The workshop agenda below shows the sequence of events for the two-day meeting.

Exhibit 9.9: Standard-Setting Workshop Agenda

September 14	PM	Table Leader Training
September 15	AM	Full-Group Training
	AM	OIB Review
	PM	Round 1
	PM	Round 2

Throughout the standard setting, the panelists had many opportunities to reflect on the pattern of performance standards they were recommending. Their general conclusion was that they were satisfied that the performance standards made sense from a content and experiential point of view. They felt the patterns reflected the requirements of the content standards and the realities of student performance.

The panelists recommended performance standards that followed an orderly progression of increasing achievement across levels. This fact is evident by examining the scale scores and percentage of students in each achievement level according to the recommended cut scores. The result from the workshop was a set of acceptable ranges for the cut scores. Specifically, Level 2 includes scale scores greater than or equal to a cut score between 404 and 408; Level 3 scale scores are greater than or equal to a cut score in the 462 to 484 range; and Level 4 scale scores begin in the 500 to 519 range.

Exhibit 9.10 shows the scale score ranges associated with cut scores recommended by the panelists, the percentage of students projected to be in each of the achievement level categories (consistently using either low, median, or high cut points for all levels), and the cumulative percentage of students at and above each achievement level.

Exhibit 9.10: Estimated Percentage of Students at Each Achievement Level

Achievement Level		Scale Score Cut Point	Percentage of Students at Level	Percentage of Student at or Above Level
Level 1	Low	--	8.70	100.00
	Median	--	9.30	100.00
	High	--	9.51	100.00
Level 2	Low	404	18.30	91.30
	Median	407	20.80	90.70
	High	408	30.25	90.49
Level 3	Low	462	23.10	73.00
	Median	468	27.67	69.90
	High	484	21.55	60.24
Level 4	Low	500	49.90	49.90
	Median	513	42.23	42.23
	High	519	38.69	38.69

References

Agresti, A. (1990). *Categorical data analysis*. New York: Wiley.

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.

American Institutes for Research. (2007). *South Carolina Alternate Assessment spring 2007 standard setting technical report.* Washington, DC: American Institutes for Research.

American Institutes for Research. (2008). South Carolina Alternate Assessment (SC-Alt): Technical report for english language arts and mathematics field test administration, spring 2006. Washington, DC: American Institutes for Research.

American Institutes for Research and South Carolina Department of Education. (2008). *South Carolina's Alternate Assessment, SC-Alt spring 2007 operational administration*. Washington, DC: American Institutes for Research.

American Institutes for Research and South Carolina Department of Education. (2009). *South Carolina's Alternate Assessment, SC-Alt spring 2008 operational administration*. Washington, DC: American Institutes for Research.

American Institutes for Research and South Carolina Department of Education. (2010). *South Carolina Alternate Assessment 2010 standard setting: Setting standards in high school biology technical report.* Washington, DC: American Institutes for Research.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*(2), 81–105.

Center for Research on Education. (2009a). Alternate assessment alignment study of replacement items report to the South Carolina State Department of Education. Columbia, SC: South Carolina State Department of Education.

Center for Research on Education. (2009b). *High school alternate assessment alignment report to the South Carolina State Department of Education*. Columbia, SC: South Carolina State Department of Education.

Cizek, G. J., & Bunch, M. B. (2007). Standard setting: A guide to establishing and evaluating performance standards on tests. Thousand Oaks, CA: Sage Publications.

Education Oversight Committee. (2008a). Review of the SC-Alternate English language arts and mathematics assessments report to the South Carolina State Department of Education. Columbia, SC: South Carolina State Department of Education.

Education Oversight Committee. (2008b). Review of the SC-Alternate science assessment report to the South Carolina State Department of Education. Columbia, SC: South Carolina State Department of Education.

Ferrara, S., Perie, M., & Johnson, E. (2008). Matching the judgmental task with standard setting panelist expertise: The Item-Descriptor (ID) Matching procedure. *Journal of Applied Testing Technology*, *9*, 1–22.

Fiske, D. W. (1995). Reprise, new themes, and steps forward. In P. E. Shrout & S. T. Fiske (Eds.), *Personality research, methods, and theory: A festschrift honoring Donald W. Fiske.* Hillsdale, NJ: Lawrence Erlbaum.

Flowers, C., Browder, D., Wakeman, S., & Karvonen, M. (2006a). *Alternate assessment alignment pilot study report to the South Carolina State Department of Education*. Columbia, SC: South Carolina State Department of Education.

Flowers, C., Browder, D., Wakeman, S., & Karvonen, M. (2006b). *Science alternate assessment alignment pilot study report to the South Carolina State Department of Education*. Columbia, SC: South Carolina State Department of Education.

Flowers, C., Browder, D., Wakeman, S., & Karvonen, M. (2007). Science alternate assessment alignment pilot study report to the South Carolina State Department of Education, Addendum, December 21, 2007. Columbia, SC: South Carolina State Department of Education.

Huynh, H. (2003, August). *Technical memorandum for computing standard error in Bookmark standard setting* [The South Carolina PACT 2003 Standard Setting Support Project]. Columbia, SC: University of South Carolina.

Kleinert, H. L., & Kearns, J. F. (2001). Alternate assessment: Measuring outcomes and supports for students with disabilities. Baltimore, MD: Brookes.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, *47*, 149–174.

National Alternate Assessment Center. (2005). *Designing from the ground floor*. Conference Pre-Session, Council of Chief State School Officers' Large-Scale Assessment Conference. San Antonio, TX: June 17–18, 2005.

Rudner, L.M., & Schafer, W.D. (2001), Reliability (ERIC Digest). College Park, MD: ERIC Clearinghouse on Assessment and Evaluation (ERIC Document Reproduction Service No. ED458213). Retrieved May 17, 2009, from www.ericdigests.org/2002-2/reliability.htm.

Sireci, S. G., Thissen, D., & Wainer, H. (1991). On the reliability of testlet-based tests. *Journal of Educational Measurement*, *28*(3), 234–247.

Spitzer, R., Cohen, J., Fleiss, J., & Endicott, J. (1967). Quantification of agreement in psychiatry diagnosis: A new approach. *Archives of General Psychiatry*, *17*, 83–87.

Appendices

Appendix A: Assignment of Tasks to Grade-Band Forms for the Spring 2010 Administration

All tasks in each SC-Alt grade-band assessment align with the Assessment Standards and Measurement Guidelines (ASMGs) in that grade-band. Because adjacent grade-band score scales are linked psychometrically, some tasks in each grade-band assessment align with ASMGs in both adjacent grade-bands. In turn, these separate grade-band ASMGs link to separate grade-level performance standards for the appropriate grades.

All items in linking tasks are developed to be appropriate for students in both adjacent grade-bands. In some cases (e.g., some tasks in ELA), the ASMGs to which linking tasks align are equivalent for two adjacent grade-bands. However, the grade-level performance standards to which the ASMGs are linked *do* differ across the adjacent grade-bands. In all content areas and for all grade-bands, Descriptions of Achievement Levels (DALs) are specific to each grade-band and differ across grade-bands.

2010 Operational Test Designs to Support Psychometric Linking of Grade-Band Score Scales

To provide data to link all grade-band assessments onto a vertical scale, linking tasks were repeated in adjacent grade-band assessments. For example, three of the tasks that appeared in the ELA grade-band 3–5 assessment also appeared in the ELA grade-band 6–8 assessment. Those three linking tasks and the nine unique tasks made up the 12 tasks in the ELA grade-band 3–5 assessment. The ELA grade-band 6–8 assessment included the three linking tasks from the grade-band 3–5 assessment, five linking tasks that also appeared in the grade-band 9–12 assessment, and four unique tasks. This "linking upward" design ensures that students were assessed on ASMGs aligned with their current grade placement or previous grades.

English Language Arts Assessment

The ELA assessment covered ASMGs in reading, writing, and communication. The 2010 operational test forms included 12 tasks for each of three grade-band assessments. Eight of these tasks were used to link between adjacent grade-band assessments, so there was a total of 28 tasks. The design for the ELA assessment for spring 2010 appears in Exhibit A-1.

Exhibit A-1: Numbers of Tasks in Each Operational Grade-Band Assessment, ELA

Grade-Band	Unique Tasks	Linking Tasks		Total for Operational Test
	Tasks in Each Grade-Band Assessment Test Booklet			
3–5	9	3	—	12
6–8	4		5	12
10	7	—		12
	Tasks to Be Included			
All grades	20	8		36

Mathematics Assessment

The mathematics assessment covered the mathematics ASMGs. The 2010 operational test forms included 12 tasks for each of three grade-band assessments, with 17 unique and eight linking tasks, for a total of 25 tasks. The design of the mathematics assessment for spring 2010 appears in Exhibit A-2.

Exhibit A-2: Numbers of Tasks in Each Operational Grade-Band Assessment, Mathematics

Grade-Band	Unique Tasks	Linking Tasks			Total for Operational Test
	Tasks in Each Grade-Band Assessment Test Booklet				
3–5	5	4	—	3	12
6–8	4		1		12
10	8	—			12
	Tasks to Be Included				
All grades	17	8			36

Science Assessment

The science assessment covered the science ASMGs. The 2010 operational test forms included 12 tasks for each of three grade-band assessments with 22 unique and seven linking tasks, for a total of 29 tasks. The design for the science assessment for spring 2010 appears in Exhibit A-3.

Exhibit A-3: Numbers of Tasks in Each Operational Grade-Band Assessment, Science

Grade-Band	Unique Tasks	Linking Tasks		Total for Operational Test
	Tasks in Each Grade-Band Assessment Test Booklet			
3–5	8	4	—	12
6–8	5		3	12
10	9	—		12
	Tasks to Be Included			
All grades	22	7		36

Social Studies Assessment

The social studies assessment covered the social studies ASMGs. The 2010 operational test forms included 12 tasks for each of two grade-band assessments, with 16 unique and four linking tasks, resulting in a total of 20 tasks. The design for the social studies assessment for spring 2010 appears in Exhibit A-4.

Exhibit A-4: Numbers of Tasks in Each Operational Grade-Band Assessment, Social Studies

Grade-Band	Unique Tasks	Linking Tasks	Total for Operational Test
	Tasks in Each Grade-Band Assessment Test Booklet		
3–5	8	4	12
6–8	8		12
	Tasks to Be Included		
All grades	16	4	24

Appendix B: Starting and Stopping Rules for Using the Student Placement Questionnaire

Directions for Determining the Starting and Concluding Tasks and Use of the Student Placement Questionnaire, Spring 2010

These directions guide you through the following steps:

- completing the Student Placement Questionnaire (SPQ),
- identifying the starting task in each content area,
- adjusting the starting task, if that becomes necessary,
- determining when to conclude the administration

Completing the Student Placement Questionnaire

The SPQ is designed to identify the most appropriate starting task for each of your students in each content area of SC-Alt. You will use the SPQ to identify the most appropriate starting task for each student in the SC-Alt assessments in English language arts, mathematics, science, and social studies. Answer each SPQ item as accurately as you can based on your experience in the classroom with this student.

The SPQs are located in the Student Answer Folder along with the areas for recording the student's scores on each SC-Alt task. An example of a completed English Language Arts SPQ is included at the end of these instructions.

Identifying the Starting Task for a Student in Each Content Area

1. Bubble in your responses to the SPQ questions.

   After you respond to all items in the SPQ, identify the most appropriate starting task for this student following the steps on the SPQ. These are the steps:

2. Count the number of bubbles you marked in each of the first three columns, and write the totals in the blocks under each column.

3. In section 3 at the bottom of the page:

   – a. Write the column totals in the appropriate blocks.

   – b. Multiply each total by the specified multiplier, and write the resulting totals in the blocks to the right.

   – c. Sum the three totals to obtain the total SPQ score. Write the SPQ score into the blocks and bubble in the SPQ score.

   – Please check your work and complete the bubble grids for the total SPQ score.

4. Find the total SPQ score in section 4 to determine the starting task for this student.

Administering the Starting Task and Completing the Administration

After you identify the starting task for this student using the SPQ, follow these directions to administer the starting task and complete the administration.

The SPQ provides the initial starting point for a student's administration. Each student must be administered a minimum of five tasks (including the starting task) if the student is started at Task 1 or a minimum of seven tasks if the student is started at Task 3 or Task 6. The minimum number of tasks and specific tasks that must be administered to each student for each starting level are specified in the table below.

Exhibit B-1: Minimum Task Ranges to Be Administered

ELA, Mathematics, Science, and Social Studies	
Starting task	**Administer all items in at least these tasks**
Task 1	1–5
Task 3	3–9
Task 6	6–12

It may be necessary to adjust the starting task based on the student's level of success on the first task. Also, the administration should be continued beyond the minimum number of tasks when the student is responding successfully.

When the Student Does Not Respond Successfully on the First Task

"Responding successfully" means getting at least three total points on a task. Each task has at least four items. Responding successfully would mean that a student received at least three total points for all the items combined. For example, a student may respond successfully by receiving three points on one item, or two points on one item and one point on another item, or one point each on three different items. When a student does not receive three or more total points on a task, the student has not responded successfully on the task.

When a student is started at Task 3 or at Task 6 and does not respond successfully on the first task, the starting task was too difficult, and the teacher must restart the student at the next lower starting point. For example:

- If the student starts at Task 3 but cannot respond successfully on Task 3, restart the student at Task 1.
- If the student starts at Task 6 but cannot respond successfully on Task 6, restart the student at Task 3.

When a student is started at Task 1, no downward adjustment is possible, and the administration must progress through at least five tasks.

When to Conclude the Administration

If the student responds successfully on the last required task as specified in the table above, continue with the administration by administering the next task and subsequent tasks until the

student no longer responds successfully on a task. By continuing the administration of subsequent tasks when the student is "responding successfully," you will provide the maximum opportunity for the student to demonstrate his or her knowledge and skills.

If the student does not respond successfully on the last required task or if at any point the student does not respond successfully on additional tasks (i.e., obtain three or more points on the task), you may conclude the administration.

By concluding the administration when the student is no longer successful after administering the required tasks, the student's test administration is not prolonged unnecessarily, and possible negative effects on the student are avoided.

Examples:

- Student A was started at Task 1 and administered Tasks 1–5. The student responded successfully on Task 5 and therefore was administered Task 6. The student responded successfully on Task 6 and was administered Task 7. The student did not respond successfully on Task 7, and the administration was concluded after Task 7.
- Student B was started at Task 3 and was administered Tasks 3–9. The student did not respond successfully on Task 9, and the administration was concluded after Task 9.

SC - ALT STUDENT PLACEMENT QUESTIONNAIRE
ENGLISH LANGUAGE ARTS

(completed SPQ example)

Follow steps 1-4 to complete the SPQ and identify the starting task.

(1) Please darken the bubble (●) that corresponds to the most appropriate response for this student. Mark only one response for each item. Please mark a response for all items below. Use a No. 2 pencil only.

	Independently	With verbal/gestural prompting	With physical prompting/hand-over-hand	No, she/he cannot do this
In reading, can this student:				
1. Attend to text read aloud?	○	●	○	○
2. Recall details of text read aloud?	○	●	○	○
3. Recognize some high-frequency written words?	○	○	●	○
4. Draw conclusions or make inferences about texts?	○	○	○	●
In writing, can this student:				
5. Write his or her name using a pencil, name stamp, letter titles, or other means?	○	●	○	○
6. Use objects, pictures, and/or picture symbols to write in any format?	○	○	●	○
7. Copy, trace, or print letters?	○	●	○	○
8. Use oral language and/or letters and words to write?	○	○	○	●
In communicating, can this student:				
9. Listen (i.e., demonstrate receptive behavior) and respond?	●	○	○	○
10. Participate in conversations by responding appropriately?	○	●	○	○
11. Use language to express a preference, opinion, or viewpoint?	○	●	○	○
12. Recognize and understand the meaning of environmental signs (e.g., street signs, store signs, school signs)?	○	●	○	○
(2) Write in the total number of bubbles you marked in each column	1	7	2	
	Col. 1	*Col. 2*	*Col. 3*	

(3) Calculate the SPQ total score

(a) write the column totals from (2) in (a) below
(b) multiply and write the results in (b) below
(c) sum the results from (b) and write the sum in (c)

	(a)		(b)
Column 1 Total	1	x 3 =	3
		+	
Column 2 Total	7	x 2 =	14
		+	
Column 3 Total	2	x 1 =	2
(c) Total SPQ Score =			19

(4) Identify the starting task for this student using the SPQ total score from step (3).

If the total score is in this range	Start at this task	Administer all items in at least these tasks
0-11	Task 1	1-5
12-22	Task 3	3-9
23-36	Task 6	6-12

DO NOT WRITE IN THIS AREA

Appendix C: Scoring Audits and Analysis of Video-Rater Data from the Spring 2009 Operational Administration

A videotaping study was conducted to audit scoring accuracy for the spring 2009 administrations of the SC-Alt in ELA, mathematics, science, and social studies. **Scoring accuracy** refers to the degree to which teachers follow scaffolding and scoring directions correctly and assign correct scores to student responses. This appendix describes the sampling procedures, the identified sample of students, and the results for the attained sample of completed videotaped administrations.

While no video-rater study was performed in 2010, the 2009 study is presented here because the results of the video-rater studies in 2008 and 2009 were rather similar. In this sense, the results of the 2009 video-rater study provide useful information about the 2010 operational SC-Alt administration.

Sampling Procedures

The sampling procedure was designed to include administrations from every school district and to be broadly representative of the range of student and test administrations. A sample of students was identified for videotaping so that (a) all districts implementing the SC-Alt would be required to videotape at least one student administration (all content areas) and (b) the total number of taped administrations per district would be based on the number of teachers involved in the assessment for each district. The sampling was by teacher and student within districts. One-third of the teachers within each district was randomly sampled to videotape the administration of one student. The number of teachers (and students) to be selected from each district is shown in Exhibit C-1.

Exhibit C-1: District-Based Sampling Targets for Video-Rater Study

Total Number of Teachers per District	Number Required to Videotape
1–5	1
6–8	2
9–11	3
12–14	4
15–17	5
18–20	6
21–23	7
24–26	8
27–29	9
30–32	10
33–35	11
36–38	12
39–41	13
42–44	14
45–47	15

Total Number of Teachers per District	Number Required to Videotape
48–50	16
51–53	17
54–56	18
57–59	19
60–62	20

Based on this sampling plan and the numbers of pre-identified students coded for each district for the 2009 administration, the frequency distribution of test administrations sampled per district was as follows:

- 1 test administration – 32 districts
- 2–5 test administrations – 26 districts
- 6–10 test administrations – 7 districts
- 11–15 test administrations – 5 districts
- 16–20 test administrations – 1 district

The sampling of students and teachers was conducted from the January 2009 precode file, which was the pre-identification file for the spring 2009 SC-Alt administration. The sampling was conducted by SCDE, and the students identified for videotaping were flagged on the precode file sent to Measurement Incorporated (MI) for the production of materials and district notification. The numbers of students by form and disability sampled for videotaping are reported in Exhibit C-2.

Exhibit C-2: Stratified Sample of Students Identified for Videotaped Administrations

	Elementary		Middle		High School		Total	
	N	%	N	%	N	%	N	%
Primary Disability								
Severe Mental Disability	16	14.3	9	11.4	7	10.9	32	12.5
Moderate Mental Disability	29	25.9	26	32.9	25	39.1	80	31.4
Mild Mental Disability	34	30.4	18	22.8	18	28.1	70	27.5
Autism	21	18.8	8	10.1	6	9.4	35	13.7
Other Disabilities	12	10.7	18	22.8	8	12.5	38	14.9
TOTAL	112		79		64		255	

Videotaping Procedures

The district test coordinators for alternate assessment were provided rosters of the students identified for videotaping. The district materials included a packet of information for each teacher that included the following:

- A videotaping student roster identifying the student

- Information on the purpose of the videotaping and instructions for how to conduct the videotaped administrations
- A videotaping student information form
- Bar code labels for positive identification and linking of the videotapes and the student information to the SC-Alt assessment data file
- Directions for the packaging and return of materials

Communications to teachers and district test coordinators emphasized the importance of completing the videotaped administrations, provided contact information for questions or concerns, and asked for notification of SCDE if there was a problem in completing a videotaped administration for a particular student. Districts notified SCDE about a small number of students who either could not be assessed (e.g., because the students had moved, the parents did not consent to videotaping, or the students were not going to be assessed with the SC-Alt) or for whom the videotaping was inappropriate or extremely difficult to implement (e.g., medical homebound students or incompatible student behavior due to taping). These students were deleted from the videotaping sample.

Approximately one-half of the students who were deleted from the videotaping sample list by SCDE were replaced by another student with the same teacher or, in a few cases, by identifying a different teacher and student. The replacement students were selected to match the grade-span form and disability of the original students as closely as possible. As a result of notifications by districts, the SCDE deleted 35 students from the original sample and instructed districts to videotape 17 replacement students.

Analysis of Video-Rater Data

The total number of students identified for videotaping after SCDE adjustments (resulting from deletions and replacements) was 237. Videotaping materials were received for 225 of the students. Of these, 12 sets of videotapes were excluded from the analysis due to nonviewable administrations. Additionally, two videotape records could not be linked to operational data. The final number of students in the attained sample was 211. This sample is summarized in Exhibit C-3.

Exhibit C-3: Demographic Frequencies for the Video-Rater Data Sample—by Test Form

	Elementary		Middle		High	
	N	%	N	%	N	%
Student's Ethnicity						
African American	49	48.51	30	49.18	29	59.18
African American/American Indian	1	0.99	.	0.00	2	4.08
American Indian	.	0.00	.	0.00	.	0.00
Asian	1	0.99	1	1.64	.	0.00
Hawaiian/Pacific Islander	.	0.00	.	0.00	.	0.00
Hispanic	4	3.96	2	3.28	.	0.00
White	44	43.56	27	44.26	18	36.73

	Elementary		Middle		High	
	N	**%**	**N**	**%**	**N**	**%**
White/African American	2	1.98	.	0.00	.	0.00
White/American Indian	.	0.00	.	0.00	.	0.00
White/Asian	.	0.00	.	0.00	.	0.00
Other	.	0.00	1	1.64	.	0.00
Student's Gender						
Female	37	36.63	16	26.23	19	38.78
Male	64	63.37	45	73.77	30	61.22
ESL (Language)						
Advanced	.	0.00	.	0.00	.	0.00
Advanced Waiver	.	0.00	.	0.00	.	0.00
Beginner	.	0.00	.	0.00	.	0.00
Beginner Waiver	.	0.00	.	0.00	.	0.00
English Speaker I	.	0.00	.	0.00	.	0.00
English Speaker II	97	96.04	59	96.72	49	100.00
Full English Proficient	.	0.00	.	0.00	.	0.00
Intermediate	.	0.00	.	0.00	.	0.00
Intermediate Waiver	.	0.00	.	0.00	.	0.00
Pre-functional	4	3.96	1	1.64	.	0.00
Pre-functional Waiver	.	0.00	1	1.64	.	0.00
Title III First Year Exited	.	0.00	.	0.00	.	0.00
Title III Second+ Year Exited	.	0.00	.	0.00	.	0.00
Eligible for Free or Reduced-Price Lunch						
Free	66	65.35	38	62.30	29	59.18
Reduced	7	6.93	1	1.64	5	10.20
No	28	27.72	22	36.07	15	30.61
EFA Grade (Reported Grade for Funding)						
1	1	0.99	.	0.00	.	0.00
2	5	4.95	.	0.00	.	0.00
3	45	44.55	.	0.00	.	0.00
4	30	29.70	1	1.64	.	0.00
5	19	18.81	14	22.95	.	0.00
6	.	0.00	16	26.23	1	2.04
7	1	0.99	14	22.95	1	2.04
8	.	0.00	11	18.03	2	4.08
9	.	0.00	2	3.28	22	44.90
10	.	0.00	3	4.92	23	46.94
11	.	0.00	.	0.00	.	0.00
12	.	0.00	.	0.00	.	0.00

	Elementary		Middle		High	
	N	**%**	**N**	**%**	**N**	**%**
Blank	.	0.00	.	0.00	.	0.00
Completion Status: Student Satisfied Attemptedness Rule						
ELA	100	99.01	61	100.00	48	97.96
Mathematics	101	100.00	61	100.00	47	95.92
Science[4]	67	66.34	46	75.41	47	95.92
Social Studies[4]	69	68.32	31	50.82	.	0.00
Completion Status: Student Did Not Answer any Content Area Items						
ELA	1	0.99	.	0.00	1	2.04
Mathematics	.	0.00	.	0.00	2	4.08
Science[4]	33	32.67	15	24.59	2	4.08
Social Studies[4]	31	30.69	30	49.18	49	100.00
Completion Status: Student Received Fewer than Five Scored Responses						
ELA	.	0.00	.	0.00	.	0.00
Mathematics	.	0.00	.	0.00	.	0.00
Science[4]	1	0.99	.	0.00	.	0.00
Social Studies[4]	1	0.99	.	0.00	.	0.00
Migrant Status	.	0.00	.	0.00	.	0.00
Home schooled	.	0.00	.	0.00	.	0.00
Medical Homebound	1	0.99	.	0.00	1	2.04
IEP Disability Codes (Multiple Codes per Student)						
Severe Mental Disability	14	13.86	6	9.84	6	12.24
Moderate Mental Disability	26	25.74	22	36.07	18	36.73
Mild Mental Disability	31	30.69	17	27.87	15	30.61
Autism	21	20.79	8	13.11	3	6.12
Deaf/Blindness	.	0.00	.	0.00	.	0.00
Emotional Disability	.	0.00	.	0.00	.	0.00
Hearing Impairment	2	1.98	1	1.64	3	6.12
Learning Disability	1	0.99	1	1.64	.	0.00
Multiple Disabilities	8	7.92	5	8.20	3	6.12
Other Health Impairment	6	5.94	5	8.20	3	6.12
Orthopedic Impairment	3	2.97	3	4.92	3	6.12
Speech Language Impairment	80	79.21	35	57.38	16	32.65
Traumatic Brain Injury	.	0.00	.	0.00	1	2.04
Visual Impairment	9	8.91	2	3.28	1	2.04
TOTAL	101	100.00	61	100.00	49	100.00

4 The completion rates for science and social studies for the elementary and middle school forms were lower due to sampling of participation in these content areas for two grade-level groups for each form (i.e., students were administered either science or social at these grade levels).

Comparing the attained video-rater (VR) sample to the identified sample (see Exhibit C-2), the following statements can be made:

By Form

- The attained sample approximates the expected number of students for each form: Elementary students make up 47.9% of the sample, middle school students make up 28.9% of the sample, and high school students make up 23.2% of the sample.

By IEP Disability Code

The first four rows show the primary disabilities of severe, moderate and mild mental disability, and autism. If any of the mental disabilities were coded together with autism, then only the mental disability is reported. Subsequent rows show additional disabilities coded by the test administrators. Since multiple disability codes per student are permitted, the percentages do not add up to 100.

- Severe mental disability was sampled similarly to the expectation across forms (elementary school: 13.9%; middle school: 9.8%; and high school: 12.2%).
- Moderate mental disability was sampled at a similar rate (25.7%, 36.1%, and 36.7%) to the expectation.
- Mild mental disability was sampled at a similar rate (30.7%, 27.9%, and 30.6%) to the expectation.
- Autism was sampled at a higher rate than expected in elementary and middle school and a lower rate in high school (20.8%, 13.1%, and 6.1%).
- The total percentages of students in primary disability categories other than severe, moderate and mild mental disability, and autism were represented at lower rates in elementary and middle school and a similar rate in high school compared with the identified sample (8.9%, 13.1%, and 14.4%).

Comparing the attained VR sample with the assessed population (see Exhibit 3.3), the following statements can be made:

By Other Demographic Variables

For other demographic variables, the proportions in the attained VR sample generally appear to correspond to those seen in the total assessed population when data were available.

- In the sample, African American (49%–59%),[5] African American/American Indian (0%–4%),Asian (0%–2%), Hispanic (0%–4%), White (37%–44%), White/African American (0%–2%), and Other (0%–2%) ethnicities were reported, representing the majority of

[5] The percentage range is reported across the three levels for which there are test forms—elementary, middle, and high school.

ethnicities in the total assessed population. These percentages evidence some variability around the corresponding population values as a result of the small sample size.

- Gender is distributed as approximately two to three males for each female; this ratio is greatest for the middle school form.
- "English Speaker II" (96% to 100%) in the sample reflects the percentage of students in the assessed population.
- Between 59% and 65% of students in the sample were eligible for free lunch, approximately the same as in the total population. A small group of students in the sample were eligible for reduced lunch (2% to 10%), which is similar to the assessed population (8%).
- None of the students in the attained VR sample were home schooled or migrant, and only two students were medically homebound; these results are comparable to the population, with reported rates of typically less than 1% for each of these demographic variables and never more than 2%.

The attained VR sample (Exhibit C-3) appears to reasonably represent the identified sample (Exhibit C-2) as well as the full population (Exhibit C-4). The demographic variables of interest are present in the attained sample data within acceptable ranges of the identified sample and the assessed population.

Item Agreement Analysis

Within each grade-band, the absolute difference between test administrator (TA) scores and AIR video-rater (VR) scores for each item was computed. Scores that do not differ between TA and VR are noted as "equal"; scores differing by +/–1 score point are noted as "adjacent"; scores differing by more than +/–1 point are flagged as "discrepant." The agreement data are summarized by content area and grade-band in Exhibit C-4, where values indicate the average percentage of items falling within each agreement category for which there were valid matched responses across TAs and VRs.

Across content areas for the elementary school form, the majority of items (90% to 95%) were shown to be scored as "equal" between the TA and VR; "adjacent" ratings were the next most prevalent outcome (5% to 9%); and "discrepant" ratings were the least prevalent result for all content areas (1% to 2%). On the middle school form, all content area show a pattern similar to the elementary form; "equal" categorizations account for the majority of ratings (94% to 97%); the "adjacent" category is the next most prevalent (3% to 5%); and "discrepant" results account for the smallest proportion of ratings (0% to 1%). Across content areas on the high school form, "equal" ratings again account for the largest proportion (94% to 95%); "adjacent" is the next most prevalent (4% to 5%); and "discrepant" is the least prevalent (1%).

Exhibit C-4: Average Item Agreement Statistics by Grade-Band and Subject

Subject	Agreement	Elementary	Middle	High
ELA	Equal	92.70	96.14	93.73
	Adjacent	6.12	3.31	5.40
	Discrepant	1.18	0.54	0.87
Math	Equal	89.72	95.06	95.15
	Adjacent	8.78	4.55	3.93
	Discrepant	1.50	0.39	0.92
Science	Equal	90.18	94.31	94.26
	Adjacent	8.64	4.41	5.00
	Discrepant	1.18	1.28	0.74
Social Studies	Equal	94.67	96.67	X
	Adjacent	4.64	2.94	X
	Discrepant	0.69	0.39	X

Classification Consistency Analysis (as distinct from scoring consistency, discussed in the previous section)

The reported performance levels for each student are derived from a scale-score-to-performance-level conversion process. Scale scores are produced based on conversions from the raw scores assigned by the TA. From these scale scores, students were assigned to one of four performance levels (i.e., Levels 1, 2, 3, or 4) within each grade-band and content area assessment. Using the VR item scores, correspondence between reported (TA) performance levels and VR performance levels was assessed according to the kappa and weighted kappa coefficients. In ELA, mathematics, science, and social studies, consistency is assessed through *weighted kappa* (Agresti, 1990; Spitzer, Cohen, Fleiss, & Endicott, 1967), which is appropriate for ordered categories:

$$\kappa_w = \frac{\sum\sum w_{ij}\pi_{ij} - \sum\sum w_{ij}\pi_{i+}\pi_{+j}}{1 - \sum\sum w_{ij}\pi_{i+}\pi_{+j}},$$

where i is the category assigned by the TA, j is the category assigned by the VR, $w_{ij} = 1-(i-j)^2/(I-1)^2$ are the weights, π_{ij} is the probability of being classified as ij, and *"+"* indicates agreement between categories. Kappa equals 0 when the agreement is that expected by chance; and kappa equals 1 when there is perfect agreement among raters.

Under the current conditions, it must be noted that not all cases included in this analysis contained complete data. Exhibit C-5 indicates the *effective sample size* ("n"; cases with information used in the content area-by-form calculation) as well as the *total count* (indicating

students assigned to the current test form both with or without data for the specified content area).

Exhibit C-5: Agreement Statistics by Subject and Grade-Band

Subject	Statistic	Elementary (n_{TOTAL} = 101)	Middle (n_{TOTAL} = 61)	High (n_{TOTAL} = 49)
ELA	n	98	61	46
	*kw**	0.940	0.974	0.979
	95%CI	0.898 - 0.983	0.943 - 1.000	0.955 - 1.000
Math	n	97	61	47
	kw	0.912	0.956	0.968
	95%CI	0.859 - 0.965	0.917 - 0.995	0.937 - 0.999
Science	n	61	43	45
	kw	0.942	0.964	1.000
	95%CI	0.892 - 0.991	0.927 – 1.000	1.000 - 1.000
Social Studies	n	64	31	
	kw	0.927	0.949	
	95%CI	0.880 - 0.974	0.897 - 1.000	

Summary

TA and VR assignments of students to performance levels typically show high levels of agreement, as weighted kappa typically ranges from 0.91 to 1.00. Further, the 95% confidence intervals show that, while sample sizes for the current calculations may be small, the agreement indices are significantly greater than chance agreement and often approach 1.00. Based on the current evidence, we can conclude that the SC-Alt was accurately scored.

Appendix D: Descriptions of Achievement Levels (DALs)

Exhibit D-1: English Language Arts Descriptions of Achievement Levels

Performance Level	ELA Achievement Level Definitions	Grades 3–5	Grades 6–8	Grade 10
1	Students performing at level 1 demonstrate emerging academic skills and competencies in reading, writing, and communication.	Students performing at level 1 should be able to • listen (as evidenced by facial expressions, gestures, or sounds) to a variety of text read aloud; • point or eye gaze to objects, pictures, or letters to complete a writing activity; • engage (using facial expressions, gestures, or sounds) in conversations focused on objects in the immediate surroundings; • listen (as evidenced by facial expressions gestures or sounds) to a speaker.	Students performing at level 1 should be able to • point or eye gaze to objects or pictures related to a variety of grade appropriate or adapted text focused on concrete concepts, read aloud; • point or eye gaze to objects, pictures, or letters to create a simple composition; • engage in conversations focused on events in the immediate surroundings as evidenced by facial expressions, gestures, or sounds; • listen to a speaker as evidenced by facial expressions, gestures, without interrupting.	Students performing at level 1 should be able to • listen to a variety of grade appropriate/adapted texts read aloud as evidenced by facial expressions, gestures, or sounds; • point or eye gaze to objects, pictures, or letters to complete more complex written products; • engage in conversations focused on objects or events outside the immediate surroundings as evidenced by facial expressions, gestures, or sounds; • listen and respond to a speaker.
2	Students performing at level 2 demonstrate foundational academic skills and competencies in reading, writing, and communication.	Students performing at level 2 should be able to • tell or show what a grade appropriate or adapted text, which contains high frequency words, is about; • identify individual words/picture symbols; • identify story elements (e.g., main idea, events, setting, and characters); • use oral and written	Students performing at level 2 should be able to • tell or show what text that requires only literal interpretation is about (using objects, pictures, or words); • read a variety of grade appropriate/adapted texts (e.g., recipes or advertisements); • identify story elements (e.g., main idea, events, setting, characters, and conflict);	Students performing at level 2 should be able to • tell or show what a grade-appropriate or adapted text that requires simple inferences is about • read a variety of texts (e.g. recipes, advertisements, schedules, and newspapers) • identify story elements (e.g., main idea, events, setting, characters, conflict, and plot); • gather meaning from graphic

Performance Level	ELA Achievement Level Definitions	Grades 3–5	Grades 6–8	Grade 10
		language to describe; • select from a list of topics to generate ideas for written communication; • listen to a speaker without interrupting; • respond appropriately in conversations .	• make connections within and between texts; • use oral and written language to explain; • select from a list of topics to generate multiple ideas for written communication; • focus attention on a speaker and listen without interrupting; • engage in conversations by answering direct questions about familiar situations; • follow oral and/or written one-step directions.	representations; • use oral and written language to explain, inform, and describe; • generate ideas for written communication; • edit own writing; • engage in conversations by answering direct questions about the immediate environment or other familiar surroundings.
3	Students performing at level 3 demonstrate increasing academic skills and competencies in reading, writing, and communication.	Students performing at level 3 should be able to • identify story elements in text (e.g., characters, settings, events, cause and effect, and problem solution); • read words and simple sentences; • generate an idea and use words, pictures, or oral language to write; • follow one-step oral or signed directions; • communicate agreement or disagreement appropriately.	Students performing at level 3 should be able to • identify and recall details in text including main idea and characters; • draw conclusions and make simple predictions and inferences about the text; • determine meaning of unfamiliar words; • generate multiple ideas by selecting from a list and use words, pictures, or oral language to write; • initiate conversation about immediate surroundings.	Students performing at level 3 should be able to • make connections with text (plot, characters, setting); • make inferences about events in text; • understand multiple meanings of words; • compare and contrast story elements from different stories; • discriminate fact from fiction; • generate an idea and use words, pictures, or oral language to write; • follow directions to complete a task; • initiate conversations about immediate surroundings or other familiar topics.
4	Students performing	Students performing at level 4	Students performing at level 4 should	Students performing at level 4 should be

Performance Level	ELA Achievement Level Definitions	Grades 3–5	Grades 6–8	Grade 10
	at level 4 demonstrate and apply academic skills and competencies in reading, writing, and communication.	should be able to • identify story elements such as the main idea and cause and effect; • draw conclusions and make predictions about text; • read and understand the main idea of a simple paragraph; • create and edit personal written products; • follow two-step oral or signed directions; • take turns appropriately during conversation or discussion.	be able to • recognize and recall details in text, including the main idea, plot, characters, and setting; • draw conclusions and make predictions and inferences about the text; • read and understand the main idea of a simple paragraph; • explain word meanings; • create and edit personal written products; • follow oral/signed or written directions; • initiate and retell conversations.	able to • recognize and recall details in text, including the main idea, plot, characters, and setting; • draw conclusions, and make predictions and inferences about the text; • read and understand the main idea of a short story; • use context clues to understand the meaning of unknown words; • make connections within and between texts and to prior knowledge, other texts, and the world; • create and edit personal written products; • use graphic representations as sources of information.

Exhibit D-2: Mathematics Descriptions of Achievement Levels

Performance Level	Mathematics Achievement Level Definitions	Grades 3–5	Grades 6–8	Grade 10
1	Students performing at level 1 demonstrate emerging academic skills and competencies in mathematics.	Students performing at level 1 should be able to • manipulate one concrete object; • observe that two geometric figures have the same attributes; • recognize attributes of objects, such as length and weight.	Students performing at level 1 should be able to • recognize the concept of one in counting objects; • recognize that two geometric figures have the same attributes; • recognize attributes of objects, such as length, weight, and size/volume.	Students performing at level 1 should be able to • recognize the concept of one more in counting objects; • match geometric figures that have the same attributes; • respond to positional concepts such as on top of or under, off-on, above and below; • match objects by one attribute such as length, weight, and size/volume.
2	Students performing at level 2 demonstrate foundational academic skills and competencies in mathematics.	Students performing at level 2 should be able to • count objects in a set; • identify objects by one attribute (color, size, shape); • classify two - and three-dimensional concrete objects according to one attribute; • recognize positional concepts (on/off); • identify measurement tools, including graphs.	Students performing at level 2 should be able to • add and subtract using concrete objects; • sort objects by one attribute (color, size, shape); • recognize and demonstrate understanding of positional concepts (on/off, below/above); • use nonstandard units to measure; • match the correct tool to a specific task (i.e. measure length, weight, time); • identify parts of a chart, graph, or table.	Students performing at level 2 should be able to • solve addition and subtraction problems; • Identify operations (+ or -); • tell which has more in a set; • identify a repeating relationship (pattern); • sort and classify objects by one attribute, (length, height, weight. volume); • use a graph or chart to gain information.
3	Students performing at	Students performing at level 3 should	Students performing at level 3 should	Students performing at level 3 should

Performance Level	Mathematics Achievement Level Definitions	Grades 3–5	Grades 6–8	Grade 10
	level 3 demonstrate increasing academic skills and competencies in mathematics.	be able to • demonstrate addition and subtraction concretely or symbolically; • count and compare objects in a set; • sort and classify objects by attribute (shape, size); • identify three-dimensional shapes (cube, sphere, cylinder); • use nonstandard units to measure; • find answers to questions in a graph.	be able to • identify the answer to one-digit addition and subtraction problems; • identify a set as having more, fewer, or the same number as another set; • extend a repeating pattern; • compare objects by attribute; • interpret information displayed in a table.	be able to • identify the process for solving an addition or a subtraction problem; • identify and use operational symbols correctly; • estimate the number of objects in a set; • add to find value of a set of coins; • describe, create, and complete a repeating pattern; • use and organize data to create charts, graphs, and tables.
4	Students performing at level 4 demonstrate and apply academic skills and competencies in mathematics.	Students performing at level 4 should be able to • demonstrate understanding of addition and subtraction; • generate a pattern using three-dimensional shapes (cube, sphere, cylinder); • compare objects by attribute (length, size); • interpret information displayed in a graph.	Students performing at level 4 should be able to • solve addition and subtraction facts without regrouping; • describe and extend a repeating pattern; • interpret information displayed in a graph; • use data to create tables.	Students performing at level 4 should be able to • identify, compare, and construct numbers; • use operation symbols (more than less than and equal to) to solve problems; • add to find the value of a set of two or more coins; • identify, describe, create, extend, and complete a repeating pattern; • describe events as more likely or less likely to occur; • use and organize data to create and interpret graphs.

Exhibit D-3: Science Descriptions of Achievement Levels

Performance Level	Science Achievement Level Definitions	Grades 3–5	Grades 6–8	Grade 10
1	Students performing at level 1 demonstrate emerging academic skills and competencies in science.	Students performing at level 1 should be able to use their senses to • observe the outcome of a simple science investigation; • sequence growth patterns; • observe and record daily weather conditions; • recognize the sun and moon and relate them to day and night; • recognize that objects move when force is applied..	Students performing at level 1 should be able to use their senses to • chose a question (how) (what if) to conduct a scientific investigation; • identify major body parts of animals; • identify the sun and moon; • recognize that objects move when force is applied and recognize speed (fast and slow); • sort by one attribute.	Students performing at level 1 should be able to use their senses to • choose questions to conduct a simple scientific investigation; • recognize that objects move when force is applied; • recognize that an object at rest does not move; • identify physical properties of matter (e.g., freezing/melting)
2	Students performing at level 2 demonstrate foundational academic skills and competencies in science.	Students performing at level 2 should be able to • generate a question to conduct a simple scientific investigation; • sort organisms by physical characteristics; • identify daily weather conditions; • recognize the pattern of day and night; • identify the position of objects such as above/below, inside, or on top; • sort materials by observable properties.	Students performing at level 2 should be able to • carry out a simple scientific investigation to answer a question; • sort and describe materials by observable properties; • sort and identify organisms by physical characteristics; • identify patterns of day and night; • recognize that an object at rest moves when force is applied.	Students performing at level 2 should be able to • carry out a simple scientific investigation related to electricity or force and motion to answer a question; • compare magnetic and non-magnetic objects; • identify the force that makes an object move; • recognize physical changes in matter; • recognize physical properties of matter.
3	Students performing at	Students performing at level 3 should be able to	Students performing at level 3 should be able to	Students performing at level 3 should be able to

Performance Level	Science Achievement Level Definitions	Grades 3–5	Grades 6–8	Grade 10
	level 3 demonstrate increasing academic skills and competencies in science.	• select appropriate tool for gathering data; • carry out a simple scientific investigation; • classify events in sequential order; • distinguish between living and nonliving things; • identify major organs of animals; • use a graph to compare daily changes in weather conditions.	• . conduct and analyze the results of a simple scientific investigation • use graphs, tables, and charts to record data and report on the results of an investigation; • compare the characteristics of living and nonliving things; • identify what plants need to grow; • use a graph or chart to compare weather conditions each season. • classify organism into major groups.	• predict the outcome of a simple investigation and compare the results with the prediction; • compare factors that affect an electromagnet; • identify electricity as a source of energy; • relate the change in force to the change in speed; • recognize the physical properties of two or more objects.
4	Students performing at level 4 demonstrate and apply academic skills and competencies in science.	Students performing at level 4 should • plan and conduct a simple scientific investigation; • identify major organs of animals and their functions; • identify living and nonliving things in terms of a food web; • identify natural resources as renewable or nonrenewable; • compare heat and light changes from season to season using a graph. • Draw simple conclusions from tables, graphs and charts	Students performing at level 4 should be able to • plan, conduct, and carry out a simple scientific investigation; • communicate simple conclusions using tables and graphs; • identify simple machines (incline plane, lever, pulley); • compare data on temperature changes over time using a graph; • use a graph to show how heat and light change from season to season; • identify sources of light.	Students performing at level 4 should be able to • plan, conduct, and analyze the results of a scientific investigation; • draw simple conclusions from distance/time graphs or tables; • demonstrate how simple machines are used to help people (inclined plane, lever, pulley, etc.); • predict the effect of the change in force on an object; • identify water as solid, steam, or liquid.

Exhibit D-4: Social Studies Descriptions of Achievement Levels

Performance Level	Social Studies Achievement Level Definitions	Grades 3–5	Grades 6–8
1	Students performing at level 1 demonstrate emerging academic skills and competencies in social studies.	Students performing at level 1 should be able to • identify self from others; • respond to a person in authority in the home or school; • follow class rules; • engage in turn-taking; • listen to information about South Carolina history.	Students performing at level 1 should be able to • identify self from others; • respond to familiar authority figures; • follow class rules; • engage in turn-taking and sharing; • listen to information presented about significant and historical events in South Carolina.
2	Students performing at level 2 demonstrate foundational skills and competencies in social studies.	Students performing at level 2 should be able to • identify characteristics such as gender that help identify self in relation to others; • match workers to different jobs in the community; • recognize people in authority and follow class rules; • match the people we honor on some national holidays (e.g., George Washington, Martin Luther King, Jr.) with the holidays; • distinguish between past and present (match jobs of the past with jobs of the present); • match significant historical figures such as Thomas Edison to their accomplishments.	Students performing at level 2 should be able to • identify surroundings (e.g., classroom, school); • match different people to their jobs in the community; • identify people in authority and follow class rules; • demonstrate understanding of rules; • identify the people we honor on some national holidays (e.g., George Washington, Martin Luther King, Jr.); • identify the purpose of money; • match changes over time to the past and present such as communication.
3	Students performing at level 3 demonstrate increasing skills and competencies in social studies.	Students performing at level 3 should be able to • understand the concept of past and present; • demonstrate respect for people in authority; • identify major symbols of the United States; • identify why we celebrate the national holidays; • recognize that when we work we earn money	Students performing at level 3 should be able to • identify members of the larger community (e.g., police officers, fire-fighters, doctors); • demonstrate understanding of consequences of not following the rules; • identify examples of good citizenship such as honesty, courage, etc.; • identify symbols of the United States (e.g., the flag, bald

		to buy things; • identify features on a map of South Carolina (river, mountain, ocean); • answer questions about significant events related to the Civil War; • identify historical figures such as Thomas Edison, Alexander Graham Bell, etc. to their accomplishments.	eagle); • demonstrate an understanding that we work to earn money and use money to buy things; • identify changes over time such as in travel, farming, etc.; • gain information from maps, charts, and graphs; • answer questions about key historical figures and significant historical events including the civil rights movement.
4	Students performing at level 4 demonstrate and apply academic skills and competencies in social studies.	Students performing at level 4 should be able to • place personal history on a time line; • identify the roles of leaders and officials in local government (e.g., principal, mayor, governor); • identify individuals who embody qualities of good citizenship; • identify examples of respect and fair treatment; • recognize that we exchange money for goods and services; • use a key to locate geographic features on a map of South Carolina; • answer questions about key concepts related to the Civil War; • answer questions about the accomplishments of key historical figures such as Thomas Edison, Alexander Graham Bell, etc.	Students performing at level 4 should be able to • place personal and family history on a time line; • identify roles of leaders and officials in local government (e.g., principal, mayor, governor); • identify examples of the qualities of courage and patriotism; • identify examples of respect and fair treatment and their opposites; • recognize how the amount of money available determines what we can buy; • gain information from maps and charts; • identify the accomplishments of Civil Rights leaders including Rosa Parks.

Exhibit D-5: High School Biology Descriptions of Achievement Levels

Performance Level	Biology Achievement Level Definitions	Grade 10
1	Students performing at level 1 demonstrate some emerging academic skills and competencies in biology.	Students performing at level 1 should be able to • Identify a possible outcome of a simple scientific investigation • Recognize tools that could be used in a simple scientific investigation • Identify a result of a simple investigation based on observations • Identify appropriate safety instruments when conducting scientific investigations

Performance Level	Biology Achievement Level Definitions	Grade 10
		• Identify things as cellular (living) • Identify food as a source of protein, carbohydrates, or fat • Identify the source of energy in a food chain • Identify the offspring of parents • Identify adaptations that allow animals to survive in their habitat • Identify living and nonliving resources in an ecosystem • Identify natural things in the environment and things made by humans
2	Students performing at level 2 demonstrate foundational academic skills in biology.	Students performing at level 2 should be able to • Identify a prediction • Identify the outcome of a simple controlled scientific investigation • Identify scientific instruments used to make observations • Interpret simple scientific data • Identify parts of a graph • Identify appropriate safety procedures when conducting scientific investigations • Recognize cellular vs non cellular (living or nonliving) things • Recognize food as protein, carbohydrate or fat • Identify the flow of energy in a simple food web • Identify parents as a source of physical traits • Identify favorable and unfavorable traits that determine species survival • Identify a phylogenetic tree as a diagram that shows ancestry of organisms • Recognize the relationships among organisms • Identify environmental changes that can effect a population • Identify human activities that affect Earth
3	Students performing at level 3 demonstrate increasing academic skills and competencies in biology.	Students performing at level 3 should be able to • Identify the hypothesis of a simple investigation • Recognize which scientific instruments are used to collect and/or record data • Organize data in a given graph/table/model • Interpret the results of a scientific data that is displayed in a graph • Identify the outcome of a simple investigation as the same/different from the original hypothesis • Identify appropriate safety procedures required when conducting a specific scientific investigation • Recall that cells are the basic unit of life

Performance Level	Biology Achievement Level Definitions	Grade 10
		• Classify things as cellular or non cellular • Illustrate that all living things are composed of cells • Identify different types of cells, tissues, and organs • Illustrate the end product of cell division • Identify what plants need for survival • Classify different foods as protein, fat, or carbohydrate • Summarize the role of protein, carbohydrates, or fat on the body • Illustrate the flow of energy in a simple food web • Identify that chromosomes contain DNA • Identify types of traits passed on from parent to offspring • Identify offspring based on dominant parent traits • Identify the structure of DNA • Identify an organism that is better adapted to a changing habitat • Identify which organisms are most closely related by using a phylogenetic tree • Identify predator/prey relationships • Explain how environmental changes can affect a population • Identify the sequence of ecological succession • Classify human activities based on their effect on Earth (beneficial or harmful)
4	Students performing at level 4 demonstrate and apply academic skills in biology.	Students performing at level 4 should be able to • Analyze the outcome of a simple investigation and compare it to the hypothesis • Select the appropriate graph for displaying simple scientific data • Use laboratory instruments and procedures in a safe manner • Recall that all cells come from other cells • Identify a nucleus, cell membrane/wall, vacuole, and chloroplast • Recall different types of cells • Illustrate that plants and animals have different cell structures • Identify different types of cells, tissues, organs, and organ systems • Classify protein, carbohydrate, or fats based on function or description of structure • Create a food web showing the flow of energy • Summarize that plants use photosynthesis to make their own food • Identify that DNA and genes pass on specific traits to offspring

Performance Level	Biology Achievement Level Definitions	Grade 10
		• Predict physical traits of offspring based on dominant or recessive physical traits of parents • Identify a dominant trait of a given species • Identify the principal of natural selection • Explain the effect of a changing habitat on a population • Explain the relationship of two organisms based on a phylogenetic tree • Identify living counterparts of extinct organisms • Classify interrelationships among organisms within ecosystems • Predict the effect of environmental changes on a population • Illustrate the changes that occur during succession • Illustrate how human activities affect the naturally occurring processes on Earth

Appendix E: Summary of Linking Design

How South Carolina Alternate Assessment Standards and Measurement Guidelines (ASMGs) Overlap across Grade Bands

Common threads across grade bands were targeted in the development of some ASMGs to promote consistent instruction across the curriculum from grade band to grade band. The difference in the essence of some standards is subtle in a number of academic standards across grade bands, and in fact some standards are exactly the same in some of the content standards.

Structure of the tasks

- Each task has four to eight items. The student responses to each item are scored from one to four points depending on demands of the response.
- Only one beginning item in the first three to five tasks in an academic content area is an engagement item. Each of the engagement items is aligned with the academic content standards through the ASMGs. The remaining items of the tasks are aligned to the academic content standards through the ASMGs at complexity levels ranging from low to high. Since every student must respond to all items in at least a minimum of five tasks, every student must respond to items that assess his or her knowledge of content and skills at the grade band to which he or she is assigned.
- Items and tasks progress upward in complexity and difficulty across the performance levels at the assigned grade band.

Structure of linking tasks

- ASMGs from adjacent grade bands were examined for common threads linked to content across the two grade bands for use in developing linking tasks.
- For the mathematics assessment, several low-complexity tasks linked across all three grade bands.
- Some items were developed specifically to link to ASMGs that were common in academic demand across grade bands. Other items were designed specifically to assess only the ASMG content for a specific grade band.

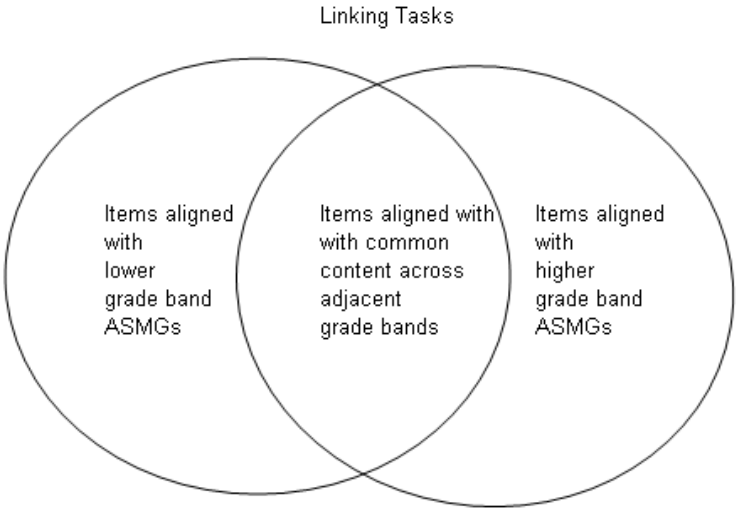

Linking Tasks
Items aligned with lower grade band ASMGs
Items aligned with with common content across adjacent grade bands
Items aligned with higher grade band ASMGs

Exhibit E-1: Summary of Linking Design Across Subjects and Grade-Bands

Subject	Grade Band	Number of Items	Number of Tasks	Starting Positions									
				Starting Task 1		Within Grade-Band Linking		Starting Task 3		Within Grade-Band Linking		Starting Task 6	
				Items	Tasks	Items	Tasks	Items	Tasks	Items	Tasks	Items	Tasks
ELA	3–5	65	12	28	5	18	3	38	7	20	4	37	7
	(Linking)	*16*											
	6–8	62	12	29	5	17	3	37	7	20	4	33	7
	(Linking)	*23*											
	10	64	12	26	5	15	3	36	7	21	4	38	7
Math	3–5	58	12	28	5	18	3	34	7	16	4	30	7
	(Linking)	*33*											
	6–8	59	12	23	5	13	3	35	7	22	4	36	7
	(Linking)	*19*											
	10	64	12	24	5	15	3	38	7	23	4	40	7
Science	3–5	60	12	23	5	13	3	34	7	21	4	37	7
	(Linking)	*19*											
	6–8	62	12	29	5	18	3	38	7	20	4	33	7
	(Linking)	*15*											
	10	56	12	26	5	15	3	33	7	18	4	30	7
Social Studies	3–5	58	13	20	6	12	4	33	8	21	5	38	8
	(Linking)	*16*											
	6–8	54	13	22	6	13	4	31	8	18	5	32	8

Appendix F: Statistics Summaries for the 2010 Spring Biology Field-Test Items

Exhibit F-1: Biology Form 1 Field-Test Classical Item Statistics

ITS Item ID	Item Position	Adjusted Biserial/ Polyserial	Average Score	Access Limitation	Omit	DIF	
						Female vs. Male	Black vs. White
1435	1	0.53	0.743590	0.00	0.02	-A	-A
1436	2	0.60	0.615385	0.00	0.03	+A	-A
1600	3	0.58	0.578947	0.01	0.05	+A	-A
1438	4	0.53	0.578947	0.01	0.04	-A	+A
1440	5	0.53	0.640351	0.01	0.05	-A	+A
1569	6	0.58	0.729345	0.00	0.01	-A	-A
1571	7	0.45	0.555556	0.00	0.03	+A	+A
1572	8	0.31	0.606838	0.00	0.03	-A	+C
1570	9	0.42	0.538462	0.00	0.06	+A	-A
1573	10	0.41	0.547009	0.00	0.03	+A	+A
1588	11	0.54	0.823349	0.00	0.02	-A	-A
1589	12	0.41	0.739535	0.00	0.03	+A	+A
1590	13	0.51	0.690141	0.01	0.03	-A	-A
1591	14	0.43	0.623256	0.00	0.03	+A	-A
1592	15	0.44	0.699531	0.01	0.04	-A	+A
1593	16	0.43	0.562791	0.00	0.06	+A	-A
1575	17	0.23	0.563981	0.01	0.05	+A	-A
1576	18	0.46	0.682464	0.01	0.04	+A	-A
1577	19	0.48	0.682464	0.01	0.04	-A	+A
1578	20	0.42	0.635071	0.01	0.04	-A	-A
1579	21	0.53	0.685714	0.01	0.04	+A	+A
1580	22	0.40	0.776190	0.01	0.03	+A	-A
1581	23	0.32	0.385714	0.01	0.07	+A	+A
1441	24	0.47	0.566038	0.00	0.06	+A	+A
1442	25	0.33	0.613208	0.00	0.03	+A	-A
1444	26	0.08	0.448113	0.00	0.05	+A	+A
1445	27	0.24	0.622642	0.00	0.05	+A	+A
1446	28	0.36	0.589623	0.00	0.05	+A	-A
1482	29	0.29	0.530661	0.02	0.03	-A	-A
1484	30	0.29	0.632075	0.02	0.03	-A	-A
1485	31	0.41	0.540095	0.02	0.03	-A	-A
1486	32	0.28	0.688679	0.02	0.01	-A	-A
1488	33	0.19	0.641509	0.02	0.03	+A	+A
1487	34	0.26	0.452830	0.02	0.05	-A	+A

ITS Item ID	Item Position	Adjusted Biserial/ Polyserial	Average Score	Access Limitation	Omit	DIF	
						Female vs. Male	Black vs. White
1551	35	0.34	0.716279	0.00	0.03	-A	+A
1552	36	0.33	0.516279	0.00	0.03	-A	+A
1553	37	0.46	0.758140	0.00	0.03	-A	+A
1554	38	0.30	0.730233	0.00	0.04	-A	-A
1555	39	0.38	0.753488	0.00	0.03	-A	+A
1556	40	0.49	0.548837	0.00	0.05	+A	-A
1546	41	0.40	0.466982	0.00	0.03	-A	-A
1545	42	0.43	0.634615	0.02	0.03	-A	+A
1548	43	0.05	0.417453	0.00	0.03	+A	-A
1547	44	0.29	0.725962	0.02	0.03	-A	-A
1549	45	0.40	0.566038	0.00	0.05	+A	+A
1550	46	0.33	0.468900	0.01	0.05	-A	-A
1498	47	0.42	0.680952	0.00	0.04	-A	-A
1500	48	0.26	0.488038	0.00	0.04	+A	+A
1502	49	0.40	0.636364	0.00	0.05	+A	+A
1505	50	0.38	0.581731	0.00	0.03	-A	+A
1506	51	0.37	0.492683	0.00	0.05	+A	+A
1507	52	0.51	0.598039	0.00	0.07	+A	+A
1508	53	0.49	0.608040	0.01	0.07	+A	-A
1458	54	0.40	0.643293	0.00	0.00	+A	+A
1460	55	0.37	0.847561	0.00	0.00	+A	-C
1459	56	0.43	0.506098	0.00	0.00	+A	+A
1461	57	0.35	0.579268	0.00	0.01	-A	+A
1463	58	0.26	0.762195	0.00	0.00	-A	+A
1462	59	0.43	0.579268	0.00	0.02	-A	+A
1470	60	0.51	0.704403	0.00	0.00	+A	+A
1472	61	0.40	0.619497	0.00	0.00	-A	+A
1473	62	0.44	0.553459	0.00	0.00	-A	+A
1474	63	0.33	0.484277	0.00	0.00	+A	+A
1475	64	0.13	0.402516	0.00	0.01	-A	-A
1434	65	0.24	0.644295	0.00	0.01	-A	-A
1447	66	0.26	0.469799	0.00	0.00	-A	-A
1448	67	0.29	0.630872	0.00	0.01	+A	+C
1449	68	0.17	0.571429	0.00	0.01	+A	-A
1450	69	0.31	0.520548	0.00	0.02	-A	-A
1451	70	0.08	0.573427	0.00	0.03	+C	-C

Exhibit F-2: Biology Form 2 Field-Test Classical Item Statistics

ITS Item ID	Item Position	Adjusted Biserial/ Polyserial	Average Score	Access Limitation	Omit	DIF	
						Female vs. Male	Black vs. White
1559	1	0.71	0.699405	0.00	0.01	+A	+A
1564	2	0.41	0.526786	0.00	0.07	+A	+A
1565	3	0.65	0.669643	0.00	0.07	+A	+A
1566	4	0.58	0.567568	0.00	0.07	+A	-A
1567	5	0.72	0.675676	0.00	0.06	+A	-A
1568	6	0.37	0.477477	0.00	0.09	+A	+A
1582	7	0.72	0.750751	0.00	0.01	+A	-A
1586	8	0.62	0.549550	0.00	0.07	-A	-A
1583	9	0.48	0.531532	0.00	0.08	-A	+A
1585	10	0.34	0.369369	0.00	0.06	-A	-A
1587	11	0.54	0.522523	0.00	0.06	-A	-A
1588	12	0.59	0.844552	0.00	0.01	+A	+A
1589	13	0.50	0.735294	0.00	0.06	+A	+A
1590	14	0.57	0.747475	0.02	0.06	+A	-A
1591	15	0.50	0.691542	0.00	0.07	-A	-A
1592	16	0.58	0.796954	0.02	0.07	-A	-A
1593	17	0.52	0.610000	0.00	0.08	+A	-A
1521	18	0.46	0.813131	0.02	0.05	+A	+A
1523	19	0.48	0.621212	0.00	0.08	+A	-A
1526	20	0.48	0.722222	0.00	0.08	-A	-A
1525	21	0.53	0.666667	0.00	0.08	-A	+A
1524	22	0.45	0.439394	0.00	0.08	+A	-A
1515	23	0.49	0.621212	0.00	0.09	+A	+A
1516	24	0.50	0.681818	0.00	0.08	-A	-A
1517	25	0.42	0.515152	0.00	0.08	-A	-A
1518	26	0.42	0.659898	0.00	0.09	-A	+A
1519	27	0.34	0.639594	0.00	0.08	-A	+A
1520	28	0.45	0.578680	0.00	0.09	+A	-A
1482	29	0.45	0.621952	0.02	0.06	-A	-A
1484	30	0.31	0.658537	0.02	0.04	+A	-A
1485	31	0.52	0.560976	0.02	0.05	+A	-A
1486	32	0.35	0.721951	0.02	0.03	-A	-A
1488	33	0.31	0.639024	0.02	0.04	-A	+A
1487	34	0.27	0.473171	0.02	0.05	+A	+A
1534	35	0.41	0.717073	0.00	0.06	+A	+A
1535	36	0.38	0.551220	0.00	0.05	-A	+C

ITS Item ID	Item Position	Adjusted Biserial/ Polyserial	Average Score	Access Limitation	Omit	DIF	
						Female vs. Male	Black vs. White
1536	37	0.40	0.721951	0.00	0.07	+A	-A
1537	38	0.39	0.712195	0.00	0.06	+A	+A
1538	39	0.49	0.697561	0.00	0.06	-A	-A
1546	40	0.19	0.420792	0.00	0.06	-A	+C
1545	41	0.24	0.595000	0.02	0.04	+A	+A
1548	42	0.04	0.420398	0.00	0.06	-A	+A
1547	43	0.37	0.720000	0.02	0.05	+A	+A
1549	44	0.49	0.597015	0.00	0.08	+A	+A
1550	45	0.24	0.460000	0.02	0.04	+A	+A
1452	46	0.48	0.620000	0.00	0.06	-A	+A
1453	47	0.44	0.633166	0.00	0.08	+A	-A
1454	48	0.45	0.732323	0.00	0.06	+A	-A
1455	49	0.30	0.489899	0.00	0.07	-A	+A
1456	50	0.47	0.376963	0.00	0.13	+A	+A
1457	51	0.41	0.704301	0.00	0.14	+A	+A
1464	52	0.44	0.697531	0.00	0.00	+A	-A
1466	53	0.37	0.820988	0.00	0.01	+A	-A
1465	54	0.43	0.666667	0.00	0.02	-A	-A
1467	55	0.23	0.518519	0.00	0.00	+A	+A
1468	56	0.06	0.689441	0.00	0.02	+A	-A
1469	57	0.43	0.677019	0.00	0.01	-A	+A
1594	58	0.52	0.825658	0.00	0.00	-A	+A
1596	59	0.53	0.697369	0.00	0.00	+A	+A
1598	60	0.46	0.805921	0.00	0.00	-A	+A
1599	61	0.53	0.845395	0.00	0.00	+A	+A
1597	62	0.35	0.750000	0.00	0.01	+A	+A
1595	63	0.45	0.815790	0.00	0.01	-A	+A
1504	64	0.39	0.723684	0.00	0.00	+A	-A
1503	65	0.39	0.688742	0.00	0.01	+A	+A
1497	66	0.53	0.725166	0.00	0.01	+A	-A
1499	67	0.43	0.768212	0.00	0.00	+A	-A
1501	68	0.24	0.600000	0.00	0.01	+A	-A

Exhibit F-3: Biology: Field-Test WINSTEPS Item Statistics

ENTRY	MEASURE	COUNT	SCORE	ERROR	IN.MSQ	IN.ZSTD	OUT.MS	OUT.ZSTD	NAME
1	-0.80972	122	285	0.116247	0.99	-0.05	1.35	1.59	its_item_1435
2	-0.48709	118	80	0.205304	0.9	-1.18	0.82	-1.6	its_item_1436
3	-0.27768	111	72	0.206405	0.98	-0.27	0.92	-0.72	its_item_1600
4	-0.30094	114	74	0.203729	0.95	-0.67	0.91	-0.85	its_item_1438
5	-0.62422	110	79	0.21912	1.01	0.13	0.96	-0.23	its_item_1440
6	-0.55856	123	281	0.104648	0.98	-0.09	1.41	1.57	its_item_1569
7	-0.19128	118	73	0.197941	1.01	0.12	0.96	-0.38	its_item_1571
8	-0.20977	119	74	0.197226	1.14	2.01	1.19	1.9	its_item_1572
9	-0.10423	112	68	0.202049	1.09	1.31	1.26	2.59	its_item_1570
10	-0.12081	118	71	0.19694	1.03	0.47	1.01	0.14	its_item_1573
11	-0.73538	439	1115	0.068403	0.93	-0.67	1.06	0.42	its_item_1588
12	-0.70466	424	328	0.12141	0.93	-1.03	0.89	-1	its_item_1589
13	-0.5133	417	312	0.11754	0.86	-2.57	0.73	-3.02	its_item_1590
14	-0.17656	423	288	0.109737	0.92	-1.79	0.89	-1.45	its_item_1591
15	-0.70083	413	322	0.123291	0.89	-1.73	0.74	-2.55	its_item_1592
16	0.127655	412	258	0.10717	0.92	-2.26	0.89	-1.73	its_item_1593
17	0.230038	212	126	0.146278	1.09	1.79	1.15	1.93	its_item_1575
18	-0.41137	214	155	0.158816	0.9	-1.34	0.84	-1.44	its_item_1576
19	-0.34178	215	153	0.156294	0.89	-1.62	0.84	-1.52	its_item_1577
20	-0.11504	215	143	0.150624	0.94	-1.06	0.9	-1.13	its_item_1578
21	-0.34753	215	153	0.156507	0.86	-2.1	0.76	-2.3	its_item_1579
22	-0.85202	216	172	0.174231	0.95	-0.47	0.87	-0.82	its_item_1580
23	1.076966	208	84	0.148309	1.02	0.37	1	0.02	its_item_1581
24	0.141747	212	130	0.147625	0.89	-2.16	0.9	-1.31	its_item_1441
25	0.022184	219	138	0.147133	1.02	0.42	0.99	-0.06	its_item_1442
26	0.752609	213	101	0.144154	1.22	4.87	1.31	4.57	its_item_1444

ENTRY	MEASURE	COUNT	SCORE	ERROR	IN.MSQ	IN.ZSTD	OUT.MS	OUT.ZSTD	NAME
27	-0.05916	215	140	0.149947	1.08	1.46	1.16	1.69	its_item_1445
28	0.101802	213	132	0.147823	0.99	-0.1	0.99	-0.11	its_item_1446
29	0.387148	419	500	0.067799	1.2	3.71	1.17	2.38	its_item_1482
30	0.006949	425	280	0.107169	1.05	1.11	1.02	0.34	its_item_1484
31	0.405936	422	480	0.077708	1	-0.02	1	-0.05	its_item_1485
32	-0.27773	431	306	0.111078	1.01	0.24	0.97	-0.29	its_item_1486
33	-0.03415	423	283	0.108027	1.08	1.94	1.14	1.92	its_item_1488
34	0.821357	417	202	0.103439	1.1	3.02	1.11	2.27	its_item_1487
35	-0.47424	220	163	0.159977	1.01	0.21	0.93	-0.58	its_item_1551
36	0.547841	220	117	0.141343	1.02	0.45	1	-0.06	its_item_1552
37	-0.74145	220	173	0.169983	0.9	-1.05	0.81	-1.39	its_item_1553
38	-0.59193	217	166	0.165751	1.04	0.49	1.04	0.33	its_item_1554
39	-0.66204	219	170	0.167338	0.99	-0.12	0.9	-0.74	its_item_1555
40	0.337441	215	125	0.144314	0.88	-2.68	0.86	-2.18	its_item_1556
41	0.864033	420	392	0.068144	1.27	4.82	1.36	5.1	its_item_1546
42	0.134314	415	264	0.107132	1.02	0.59	1.04	0.61	its_item_1545
43	0.995938	421	360	0.065606	1.75	9.9	1.95	9.9	its_item_1548
44	-0.45301	415	311	0.117682	1.02	0.4	0.95	-0.52	its_item_1547
45	0.403209	412	497	0.065846	1.14	2.66	1.2	2.4	its_item_1549
46	0.858048	412	198	0.103981	1.08	2.59	1.09	1.93	its_item_1550
47	-0.29062	214	152	0.156871	0.92	-1.09	0.88	-1.08	its_item_1498
48	0.633642	213	110	0.143569	1.04	1.01	1.21	3.23	its_item_1500
49	-0.07209	212	142	0.15212	0.94	-1.01	0.93	-0.77	its_item_1502
50	0.221037	215	130	0.146087	0.96	-0.88	0.94	-0.84	its_item_1505
51	0.694714	210	106	0.144239	0.97	-0.72	0.96	-0.6	its_item_1506
52	0.156786	206	129	0.150061	0.86	-2.78	0.81	-2.44	its_item_1507
53	0.178661	204	127	0.150359	0.87	-2.6	0.83	-2.19	its_item_1508
54	0.274028	171	220	0.111592	1	0.06	0.99	-0.04	its_item_1458

ENTRY	MEASURE	COUNT	SCORE	ERROR	IN.MSQ	IN.ZSTD	OUT.MS	OUT.ZSTD	NAME
55	-1.01616	171	146	0.219414	0.92	-0.52	0.75	-1.19	its_item_1460
56	0.847458	171	86	0.15864	0.91	-2.22	0.89	-1.81	its_item_1459
57	0.448959	170	101	0.161545	0.96	-0.86	0.94	-0.77	its_item_1461
58	-0.37737	171	130	0.183279	1	0.01	1.06	0.48	its_item_1463
59	0.454774	167	99	0.162882	0.9	-2.04	0.88	-1.53	its_item_1462
60	-0.11846	164	118	0.178205	0.83	-2.29	0.77	-2.03	its_item_1470
61	0.430165	164	207	0.104673	0.99	-0.1	0.96	-0.28	its_item_1472
62	0.645299	164	184	0.110526	0.96	-0.41	0.95	-0.55	its_item_1473
63	0.891901	164	82	0.161792	0.99	-0.32	0.99	-0.16	its_item_1474
64	1.292972	162	66	0.16594	1.11	2.09	1.12	1.7	its_item_1475
65	0.270839	153	99	0.174011	1.04	0.63	1	0.03	its_item_1434
66	1.063472	154	72	0.167241	1.03	0.67	1.02	0.36	its_item_1447
67	0.270839	153	99	0.174011	0.99	-0.19	0.95	-0.46	its_item_1448
68	0.557433	152	89	0.169878	1.07	1.34	1.11	1.41	its_item_1449
69	0.761687	151	81	0.169113	0.99	-0.25	0.96	-0.54	its_item_1450
70	0.666796	149	84	0.170514	1.13	2.66	1.15	2	its_item_1451
71	-0.57182	117	251	0.106748	0.69	-2.41	0.67	-1.74	its_item_1559
72	-0.35337	103	65	0.214915	1.09	1.2	1.08	0.76	its_item_1564
73	-1.09276	103	80	0.244392	0.92	-0.55	0.83	-0.86	its_item_1565
74	-0.53469	103	69	0.220286	0.96	-0.45	0.89	-0.83	its_item_1566
75	-0.97925	105	79	0.237344	0.85	-1.3	0.78	-1.3	its_item_1567
76	-0.02614	98	56	0.214917	1.2	2.81	1.29	2.86	its_item_1568
77	-0.8358	117	266	0.117318	0.69	-2.27	0.6	-2.2	its_item_1582
78	-0.38194	103	66	0.21612	0.94	-0.76	0.89	-0.98	its_item_1586
79	-0.31536	101	64	0.216332	1.08	1.01	1.04	0.38	its_item_1583
80	0.644498	104	43	0.209281	1.13	1.83	1.2	2.14	its_item_1585
81	-0.16645	104	62	0.210026	0.99	-0.09	0.97	-0.33	its_item_1587
82	-1.07491	202	169	0.197015	0.93	-0.58	0.76	-1.17	its_item_1521

ENTRY	MEASURE	COUNT	SCORE	ERROR	IN.MSQ	IN.ZSTD	OUT.MS	OUT.ZSTD	NAME
83	0.083208	199	129	0.156863	0.93	-1.13	0.85	-1.54	its_item_1523
84	-0.49689	199	150	0.172475	0.94	-0.74	0.81	-1.29	its_item_1526
85	-0.16119	199	138	0.162353	0.88	-1.82	0.77	-2.01	its_item_1525
86	1.00915	199	89	0.152466	0.92	-1.55	0.91	-1.29	its_item_1524
87	0.054134	197	129	0.158506	0.92	-1.29	0.85	-1.48	its_item_1515
88	-0.2228	199	140	0.163946	0.92	-1.08	0.86	-1.12	its_item_1516
89	0.638208	199	105	0.151813	0.97	-0.52	0.94	-0.71	its_item_1517
90	-0.17647	197	138	0.16365	0.99	-0.08	0.94	-0.5	its_item_1518
91	0.070172	199	129	0.157321	1.05	0.86	1.01	0.17	its_item_1519
92	0.269698	198	120	0.154804	0.94	-1.02	0.88	-1.33	its_item_1520
93	-0.40068	206	153	0.166971	0.99	-0.09	0.92	-0.58	its_item_1534
94	0.434906	210	120	0.148761	1.02	0.39	0.98	-0.21	its_item_1535
95	-0.4204	204	153	0.168754	1.01	0.11	0.99	-0.04	its_item_1536
96	-0.37062	206	152	0.165987	1	0.01	0.96	-0.28	its_item_1537
97	-0.28442	206	149	0.16323	0.9	-1.4	0.82	-1.45	its_item_1538
98	0.169386	202	129	0.154791	0.92	-1.37	0.86	-1.52	its_item_1452
99	0.051	199	133	0.158128	0.99	-0.07	0.93	-0.61	its_item_1453
100	-0.46345	202	152	0.171281	0.93	-0.81	0.83	-1.17	its_item_1454
101	0.756631	201	103	0.150589	1.06	1.19	1.17	2.18	its_item_1455
102	1.297624	186	76	0.159548	0.87	-2.33	0.83	-2.2	its_item_1456
103	-0.28468	185	137	0.175128	0.98	-0.19	0.9	-0.68	its_item_1457
104	-0.03297	167	119	0.178698	0.9	-1.25	0.86	-1.09	its_item_1464
105	-0.7856	165	138	0.215616	0.96	-0.29	0.84	-0.71	its_item_1466
106	0.150111	164	112	0.175352	0.93	-0.93	0.9	-0.84	its_item_1465
107	0.81051	168	90	0.164173	1.11	1.98	1.14	1.7	its_item_1467
108	0.023652	164	116	0.178887	1.19	2.37	1.67	4.34	its_item_1468
109	0.06275	165	115	0.177246	0.91	-1.25	0.88	-0.92	its_item_1469
110	-0.49093	158	261	0.143637	0.81	-1.48	0.69	-1.6	its_item_1594

ENTRY	MEASURE	COUNT	SCORE	ERROR	IN.MSQ	IN.ZSTD	OUT.MS	OUT.ZSTD	NAME
111	0.070862	158	222	0.127458	0.9	-1.02	0.9	-0.75	its_item_1596
112	-0.29542	157	256	0.134388	0.96	-0.28	0.93	-0.24	its_item_1598
113	-0.6311	158	267	0.150745	0.8	-1.41	0.61	-2	its_item_1599
114	-0.05758	156	233	0.128975	1.13	1.15	1.39	2.36	its_item_1597
115	-0.35058	155	258	0.138037	0.88	-0.82	0.87	-0.5	its_item_1595
116	-0.02587	157	229	0.129082	1.11	1.07	1.11	0.82	its_item_1504
117	0.047086	155	110	0.184739	0.97	-0.41	0.89	-0.81	its_item_1503
118	0.195454	156	226	0.115629	0.87	-1.32	0.82	-1.03	its_item_1497
119	-0.37523	157	122	0.199717	0.89	-1	0.76	-1.44	its_item_1499
120	0.59481	156	93	0.172339	1.1	1.71	1.16	1.71	its_item_1501

Appendix G: Marginal Reliability by Grade Band, Subject, and Starting Task

Exhibit G-1: Marginal Reliability by Starting Task and Grade Band for ELA

Initial Task	Statistic	Elementary	Middle	High
1	N	353	267	97
	Reliability	0.851	0.856	0.860
	$\bar{\sigma}_{e^*}$	20.226	22.254	24.268
3	N	326	228	53
	Reliability	0.875	0.903	0.879
	$\bar{\sigma}_{e^*}$	11.318	10.593	16.643
6	N	703	682	203
	Reliability	0.822	0.781	0.835
	$\bar{\sigma}_{e^*}$	16.433	27.408	20.133

Exhibit G-2: Marginal Reliability by Starting Task and Grade Band for Mathematics

Initial Task	Statistic	Elementary	Middle	High
1	N	361	250	95
	Reliability	0.851	0.791	0.827
	$\bar{\sigma}_{e^*}$	21.745	25.707	27.878
3	N	356	265	62
	Reliability	0.806	0.862	0.867
	$\bar{\sigma}_{e^*}$	11.262	11.152	9.917
6	N	660	661	194
	Reliability	0.743	0.808	0.861
	$\bar{\sigma}_{e^*}$	24.740	23.272	15.650

Exhibit G-3: Marginal Reliability by Starting Task and Grade Band for Science

Initial Task	Statistic	Elementary	Middle	High
1	N	295	192	106
	Reliability	0.823	0.824	0.827
	$\bar{\sigma}_{e^*}$	23.635	26.783	31.939
3	N	264	170	61
	Reliability	0.767	0.855	0.810
	$\bar{\sigma}_{e^*}$	12.615	14.569	14.513
6	N	383	440	181
	Reliability	0.718	0.745	0.768
	$\bar{\sigma}_{e^*}$	27.559	27.603	27.147

Exhibit G-4: Marginal Reliability by Starting Task and Grade Band for Social Studies

Initial Task	Statistic	Elementary	Middle	High
1	N	233	181	X
	Reliability	0.802	0.820	X
	$\bar{\sigma}_{e^*}$	31.082	29.991	X
3	N	193	129	X
	Reliability	0.866	0.821	X
	$\bar{\sigma}_{e^*}$	15.365	18.088	X
6	N	518	501	X
	Reliability	0.803	0.715	X
	$\bar{\sigma}_{e^*}$	20.467	34.264	X

Appendix H: Score Report Sample

INDIVIDUAL STUDENT REPORT

Prepared Especially for the Family of
Kyree Adams

Date of Birth:	9/20/1999
Student ID:	587412589457
School District:	Calvert
School:	Alfonso Elementary School

Spring 2010

The South Carolina Alternate Assessment (SC-Alt)

Kyree participated in the South Carolina Alternate Assessment (SC-Alt) during the spring of 2010.

She took the elementary school form of the test, which is based on academic standards from grades 3 to 5. This report is designed to provide you with information on your child's performance on this assessment.

The SC-Alt is a test designed for students with significant cognitive disabilities who participate in a school curriculum that includes academic and functional skill instruction. The alternate assessment only tests students' achievement in English language arts (ELA), mathematics, science, and social studies. Individualized Education Program (IEP) reports and other methods provide parents with information on how students are progressing in the other areas.

What is the SC-Alt?

- The SC-Alt assessment includes performance tasks in each subject area. Students may complete the tasks by using their usual method of communication. This may include pointing or gazing at answer choices, selecting objects, pictures, or picture symbols that represent an answer choice, or reading letters, words or sentences to complete the task.
- The tasks are linked to the state academic content standards in four areas: English language arts (ELA), mathematics, science, and social studies.
- Students are assigned a test form based on their age. Students ages 8-10 are assigned to the elementary school form; students ages 11-13 are assigned to the middle school form; and students age 15 take the high school form.

How are scores reported and used?

- Four achievement levels (Level 1, Level 2, Level 3, and Level 4) have been established for the SC-Alt. Achievement levels describe how students are doing in relation to the state academic standards. Your child's performance is also reported as a scale score that allows parents to monitor growth from year to year.
- Level 2 is the achievement level reported as meeting the "Basic" reporting requirement for state accountability on the District Report Card. Achievement Levels 3 and 4 are the achievement levels reported as "proficient" for schools and districts in the federal accountability Adequate Yearly Progress (AYP) report.

Where can I get more information about SC-Alt and my child's performance?

- You can contact your child's teacher or school for more information.
- You can view examples of tasks, information about expectations at each achievement level, and scale score tables on the South Carolina Department of Education website at http://www.ed.sc.gov/agency/Accountability/Assessment/SouthCarolinaAlternateAssessment.html.

The South Carolina Department of Education

Spring 2010
Kyree Adams

The SC-Alt

Mathematics

Kyree scored at Level 3 with a scale score of 480 in mathematics.

Students who score at Level 3 should be able to:

- add and subtract simple numbers;
- count and compare objects in a group;
- compare objects by color, size, or shape;
- identify three-dimensional shapes;
- read information in a graph.

	4	Students performing at Level **4** demonstrate and apply academic skills and competencies in mathematics.
Your Child's Level	3	Students performing at Level **3** demonstrate increasing academic skills and competencies in mathematics.
	2	Students performing at Level **2** demonstrate foundational academic skills and competencies in mathematics.
	1	Students performing at Level **1** may demonstrate emerging academic skills and competencies in mathematics.

How you can support Kyree's learning

- Help your child add and subtract during everyday activities. For example, show her five pennies or other objects and give her one more. Then, ask her to tell you how many there are altogether.
- Play games with your child. Use games that require matching numbers (dots), such as dominoes.
- Put three objects in one pile and two objects in another pile. Ask your child which pile has more objects.
- Describe everyday household objects by shapes. For example, a can is a cylinder; a box is a rectangular prism; and a ball is a sphere.

English Language Arts

Kyree scored at Level 4 with a scale score of 495 in ELA.

Students who score at Level 4 should be able to:

- identify the main idea and make predictions about what will happen next in a story;
- write a simple story;
- follow two-step directions;
- take turns appropriately during conversations.

Your Child's Level	4	Students performing at Level **4** demonstrate and apply academic skills and competencies in reading, writing, and research.
	3	Students performing at Level **3** demonstrate increasing academic skills and competencies in reading, writing, and research.
	2	Students performing at Level **2** demonstrate foundational academic skills and competencies in reading, writing, and research.
	1	Students performing at Level **1** may demonstrate emerging academic skills and competencies in reading, writing, and research.

How you can support Kyree's learning

- Encourage your child to read passages from a variety of materials (books, magazines, newspapers).
- Read a story with your child and talk to her about specific characters and the order of events in the story.
- Assist your child with writing about an event or activity using her typical method of communication.
- Encourage your child to begin conversations with family members or friends by using her typical method of communication.

2

Spring 2010
Kyree Adams

Science

Kyree scored at Level 1 with a scale score of 270 in science.

Students who score at Level 1 should be able to:

- use senses to observe the outcome of a simple scientific investigation;
- sequence growth patterns;
- observe and record daily weather conditions;
- recognize the sun and moon and relate them to day and night;
- recognize that objects move when force is applied.

4	Students performing at Level **4** demonstrate and apply academic skills and competencies in science.
3	Students performing at Level **3** demonstrate increasing academic skills and competencies in science.
2	Students performing at Level **2** demonstrate foundational academic skills and competencies in science.
Your Child's Level → 1	Students performing at Level **1** may demonstrate emerging academic skills and competencies in science.

How you can support Kyree's learning

- Help your child plant a seed and watch it grow.
- Explain and help your child observe different weather conditions.
- Point out the moon in the sky at night.
- Show your child two balls. Roll one across a table and leave one still. Point out which ball is in motion and which is not.

Social Studies

Kyree scored at Level 3 with a scale score of 500 in Social Studies.

Students who score at Level 3 should be able to:

- understand the concept of past and present;
- demonstrate respect for people of authority;
- identify major symbols of the United States (the flag, bald eagle);
- recognize that when we work we earn money to buy things;
- identify features on a map of South Carolina (river, mountain, ocean);
- match historical figures such as Abraham Lincoln, Thomas Jefferson, etc., to their accomplishments.

4	Students performing at Level **4** demonstrate and apply academic skills and competencies in social studies.
Your Child's Level → 3	Students performing at Level **3** demonstrate increasing academic skills and competencies in social studies.
2	Students performing at Level **2** demonstrate foundational academic skills and competencies in social studies.
1	Students performing at Level **1** may demonstrate emerging academic skills and competencies in social studies.

How you can support Kyree's learning

- Play a "Now or Long Ago" game (e.g., do we ride in a stagecoach now or long ago?).
- Find and identify items in the community such as the American flag.
- Talk about national holidays (Fourth of July and what it means) and celebrate them with your child.
- Look at a map of South Carolina with your child and find what is near the mountains and near the ocean.
- Let your child earn an allowance by doing chores or helping out by following rules. (A job can be as simple as not throwing a tantrum or allowing you to brush her teeth without fussing.)

3

Spring 2010
Kyree Adams

The SC-Alt

The following areas are tested in Mathematics:

Number and Operations
- whole numbers
- fractions
- addition and subtraction
- multiplication and division

Algebra
- patterns and their relationships

Geometry
- attributes of objects such as shape, size, color
- identification of two- and three-dimensional shapes

Measurement
- money
- length, liquid, volume, and mass and weight
- time
- equivalences

Data Analysis and Probability
- data collection and representation
- data analysis
- probability

The following areas are tested in English Language Arts:

Reading
- reading
- comprehending a variety of texts (such as fiction, nonfiction, poetry, and drama)
  Note: Reading materials may include objects, pictures or photographs, picture symbols, letters, and words.

Writing
- developing written communications (notes, stories) using the student's typical method of communication

Research
- accessing and using information from a variety of sources
- communicating their own ideas and ideas of others*
  **Students' typical method of communication, verbal or nonverbal, may be facilitated by using objects, pictures or photographs, picture symbols, letters and words, voice output devices, or assistive technology.*

The following areas are tested in Science:

Scientific Inquiry involves studying scientific processes and skills such as:
- observing
- classifying
- predicting what will happen in a simple scientific experiment

Life Science
- basic needs of plants and animals
- their structures and habitats

Earth Science
- weather
- objects in the sky (sun and moon)
- earth materials (rocks and soil)

Physical Science
- characteristics of objects
- the effect of force on the motion of objects
- light, heat, and electricity

The following areas are tested in Social Studies:

Social Studies Literacy Elements are concepts required for understanding this subject such as:
- distinguishing between past, present, and future
- demonstrating responsible citizenship within the school community, the local community, and national communities
- creating and using timelines
- understanding the relationship between people and the land

Academic Standards include concepts related to specific historical time frames:
- history
- geography
- political science/government
- economics

4

Appendix I: Student Performance by Demographics, Grade-Band, and Subject Area

Exhibit I-1: Performance by Grade-Band and Demographics—ELA[6]

	Elementary School (ES)					Middle School (MS)					High School (HS)				
	N	Scale Score		Ach. Level		N	Scale Score		Ach. Level		N	Scale Score		Ach. Level	
		Mean	SD	< 3	>= 3		Mean	SD	< 3	>= 3		Mean	SD	< 3	>= 3
Student's Age															
8	474	489	55	119	355	.	.	.	.	.	.	.	.	.	.
9	472	496	62	123	349	.	.	.	.	.	.	.	.	.	.
10	444	505	66	85	359	3	–	–	–	–	.	.	.	.	.
11	1	–	–	–	–	433	509	67	122	311	.	.	.	.	.
12	.	.	.	.	.	400	515	80	107	293	.	.	.	.	.
13	.	.	.	.	.	348	513	81	113	235	.	.	.	.	.
14	.	.	.	.	.	.	.	.	.	.	3	–	–	–	–
15	.	.	.	.	.	.	.	.	.	.	346	512	78	103	243
16	.	.	.	.	.	.	.	.	.	.	8	–	–	–	–
Student's Ethnicity															
African American	671	503	62	138	533	618	514	77	186	432	190	514	82	51	139
African American/American Indian	8	–	–	–	–	4	–	–	–	–	.	.	.	.	.
American Indian	5	–	–	–	–	3	–	–	–	–	.	.	.	.	.
Asian	11	497	43	3	8	14	505	68	4	10	4	–	–	–	–
Hawaiian/Pacific Islander	1	–	–	–	–	.	.	.	.	.	.	.	.	.	.
Hispanic	65	480	60	21	44	43	495	78	11	32	10	497	28	2	8
White	592	493	59	147	445	486	511	74	135	351	150	510	75	52	98
White/African American	22	494	58	7	15	11	502	101	3	8	3	–	–	–	–
White/American Indian	1	–	–	–	–	1	–	–	–	–	.	.	.	.	.
White/Asian	6	–	–	–	–	1	–	–	–	–	.	.	.	.	.
Other	9	–	–	–	–	3	–	–	–	–	.	.	.	.	.
Unknown	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Student's Gender															
Female	455	490	63	121	334	391	509	79	114	277	123	521	76	32	91
Male	936	500	60	207	729	793	513	74	230	563	234	508	78	75	159
Unknown	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ESL (Language)															
Pre-functional	49	476	65	16	33	31	493	69	9	22	7	–	–	–	–
Beginner	1	–	–	–	–	1	–	–	–	–	.	.	.	.	.

[6] Note: Data marked '–' are suppressed because the subgroup contains fewer than 10 students.

	Elementary School (ES)					Middle School (MS)					High School (HS)				
	N	Scale Score		Ach. Level		N	Scale Score		Ach. Level		N	Scale Score		Ach. Level	
		Mean	SD	< 3	>= 3		Mean	SD	< 3	>= 3		Mean	SD	< 3	>= 3
Intermediate	.	.	.	.	.	2	–	–	–	–	.	.	.	.	.
Advanced	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Full English Proficient	1	–	–	–	–	1	–	–	–	–	1	–	–	–	–
Title III First Year Exited	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Title III Second+ Year Exited	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
English Speaker I	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
English Speaker II	1316	497	61	309	1007	1135	512	76	333	802	348	512	78	105	243
Pre-functional Waiver	2	–	–	–	–	1	–	–	–	–	.	.	.	.	.
Beginner Waiver	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Intermediate Waiver	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Advanced Waiver	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Unknown	22	530	41	2	20	13	559	84	2	11	1	–	–	–	–
Eligible for Free or Reduced-Price Lunch															
Free Meals	892	504	63	176	716	723	521	78	190	533	212	515	78	62	150
Full-Pay Meals	390	479	55	128	262	354	497	70	121	233	113	506	68	37	76
Reduced Meals	109	502	50	24	85	107	501	72	33	74	31	509	102	8	23
Unknown	.	.	.	.	.	.	.	.	.	.	1	–	–	–	–
IEP Disability Codes (Multiple Codes per Student)															
Severe Mental Disability	146	408	67	121	25	135	416	73	116	19	38	400	85	34	4
Moderate Mental Disability	338	486	43	78	260	409	501	45	115	294	161	518	49	38	123
Mild Mental Disability	387	534	43	15	372	274	574	66	8	266	64	579	60	3	61
Autism	321	494	47	73	248	234	504	60	76	158	57	488	61	23	34
Deaf/Blindness	11	454	92	6	5	9	–	–	–	–	5	–	–	–	–
Emotional Disability	9	–	–	–	–	4	–	–	–	–	2	–	–	–	–
Hearing Impaired	31	473	50	11	20	15	532	57	3	12	5	–	–	–	–
Learning Disability	24	557	31	.	24	17	597	80	.	17	2	–	–	–	–
Multiple-Disability	109	452	70	56	53	77	467	79	51	26	17	477	80	9	8
Other Health Impaired	84	509	55	12	72	61	521	82	16	45	19	505	87	6	13
Orthopedically Impaired	85	473	63	39	46	51	511	66	14	37	13	509	121	3	10

	Elementary School (ES)					Middle School (MS)					High School (HS)				
	N	**Scale Score**		**Ach. Level**		**N**	**Scale Score**		**Ach. Level**		**N**	**Scale Score**		**Ach. Level**	
		Mean	**SD**	**< 3**	**>= 3**		**Mean**	**SD**	**< 3**	**>= 3**		**Mean**	**SD**	**< 3**	**>= 3**
Speech or Language Impaired	1037	500	53	212	825	606	508	66	177	429	124	511	62	38	86
Traumatic Brain Injury	9	–	–	–	–	6	–	–	–	–	1	–	–	–	–
Visually Impaired	49	434	83	32	17	47	477	81	23	24	14	446	94	8	6
Total	**1391**	**497**	**61**	**328**	**1063**	**1184**	**512**	**76**	**344**	**840**	**357**	**512**	**77**	**107**	**250**

Exhibit I-2: Performance by Grade-Band and Demographics—Mathematics[7]

	Elementary School (ES)					Middle School (MS)					High School (HS)				
	N	Scale Score		Ach. Level		N	Scale Score		Ach. Level		N	Scale Score		Ach. Level	
		Mean	SD	< 3	>= 3		Mean	SD	< 3	>= 3		Mean	SD	< 3	>= 3
Student's Age															
8	473	492	63	171	301	.	.	.	.	.	.	.	.		
9	470	503	71	145	325	.	.	.	.	.	.	.	.	.	.
10	443	509	73	116	327	3	–	–	–	–	.	.	.	.	.
11	1	–	–	–	–	433	507	66	151	282	.	.	.	.	.
12	.	.	.	.	.	400	511	80	141	259	.	.	.	.	.
13	.	.	.	.	.	348	507	80	128	219	.	.	.	.	.
14	.	.	.	.	.	.	.	.	.	.	3	–	–	–	–
15	.	.	.	.	.	1	–	–	–	–	344	495	72	140	204
16	.	.	.	.	.	.	.	.	.	.	8	–	–	–	–
Student's Ethnicity															
African American	668	510	69	169	498	618	511	77	213	404	190	496	80	80	110
African American/American Indian	8	–	–	–	–	4	–	–	–	–	.	.	.	.	.
American Indian	5	–	–	–	–	3	–	–	–	–	.	.	.	.	.
Asian	11	514	48	3	8	14	494	75	6	8	4	–	–	–	–
Hawaiian/Pacific Islander	1	–	–	–	–	.	.	.	.	.	.	.	.	.	.
Hispanic	65	483	69	31	34	43	490	72	17	26	10	487	59	5	5
White	591	495	69	210	381	487	507	71	178	309	148	493	63	61	87
White/African American	22	499	53	6	16	11	512	104	4	7	3	–	–	–	–
White/American Indian	1	–	–	–	–	1	–	–	–	–	.	.	.	.	.
White/Asian	6	–	–	–	–	1	–	–	–	–	.	.	.	.	.
Other	9	–	–	–	–	3	–	–	–	–	.	.	.	.	.
Unknown	.	.	.	.	.	.	.	.	.	.	.	.	.		
Student's Gender															
Female	452	494	71	155	297	391	505	80	143	248	122	500	67	48	74
Male	935	505	68	278	656	794	510	72	280	513	233	492	74	99	134
Unknown	.	.	.	.	.	.	.	.	.	.	.	.	.		
ESL (Language)															
Pre-functional	49	480	78	20	29	31	485	67	16	15	7	–	–	–	–
Beginner	1	–	–	–	–	1	–	–	–	–	.	.	.	.	.
Intermediate	.	.	.	.	.	2	–	–	–	–	.	.	.	.	.

[7] Note: Data marked '–' are suppressed because the subgroup contains fewer than 10 students.

	Elementary School (ES)					Middle School (MS)					High School (HS)				
	N	Scale Score		Ach. Level		N	Scale Score		Ach. Level		N	Scale Score		Ach. Level	
		Mean	SD	< 3	>= 3		Mean	SD	< 3	>= 3		Mean	SD	< 3	>= 3
Advanced	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Full English Proficient	1	–	–	–	–	1	–	–	–	–	1	–	–	–	–
Title III First Year Exited	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Title III Second+ Year Exited	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
English Speaker I	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
English Speaker II	1312	501	68	406	905	1136	508	75	406	730	346	495	73	141	205
Pre-functional Waiver	2	–	–	–	–	1	–	–	–	–	.	.	.	.	.
Beginner Waiver	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Intermediate Waiver	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Advanced Waiver	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Unknown	22	547	74	4	18	13	565	94	1	11	1	–	–	–	–
Eligible for Free or Reduced-Price Lunch															
Free Meals	892	510	71	226	666	723	516	77	228	494	212	497	76	88	124
Full-Pay Meals	386	478	64	176	209	355	494	72	155	200	111	488	58	50	61
Reduced Meals	109	504	53	31	78	107	501	68	40	67	31	500	82	9	22
Unknown	.	.	.	.	.	.	.	.	.	.	1	–	–	–	–
IEP Disability Codes (Multiple Codes per Student)															
Severe Mental Disability	147	405	69	130	17	135	411	72	125	10	38	388	70	37	1
Moderate Mental Disability	338	487	45	124	214	410	497	46	162	248	160	500	45	62	98
Mild Mental Disability	385	544	53	27	358	274	569	61	15	259	64	554	53	7	57
Autism	319	497	52	104	214	234	504	60	84	149	56	483	64	28	28
Deaf/Blindness	11	476	110	6	5	9	–	–	–	–	5	–	–	–	–
Emotional Disability	9	–	–	–	–	4	–	–	–	–	2	–	–	–	–
Hearing Impaired	31	478	56	10	21	15	542	76	4	11	5	–	–	–	–
Learning Disability	24	574	55	.	24	17	580	65	1	16	1	–	–	–	–
Multiple-Disability	110	456	73	67	43	78	463	81	52	26	16	464	73	10	6
Other Health Impaired	84	513	69	20	64	62	519	81	21	41	18	481	75	9	9
Orthopedically Impaired	85	473	71	44	41	51	502	67	18	33	13	478	108	6	7

	Elementary School (ES)					Middle School (MS)					High School (HS)				
	N	**Scale Score**		**Ach. Level**		**N**	**Scale Score**		**Ach. Level**		**N**	**Scale Score**		**Ach. Level**	
		Mean	**SD**	**< 3**	**>= 3**		**Mean**	**SD**	**< 3**	**>= 3**		**Mean**	**SD**	**< 3**	**>= 3**
Speech or Language Impaired	1033	504	59	298	734	606	505	67	217	389	124	501	60	51	73
Traumatic Brain Injury	8	–	–	–	–	6	–	–	–	–	1	–	–	–	–
Visually Impaired	50	436	93	34	16	47	470	77	27	20	14	428	90	8	6
TOTAL	**1387**	**501**	**69**	**433**	**953**	**1185**	**508**	**75**	**423**	**761**	**355**	**495**	**72**	**147**	**208**

Exhibit I-3: Performance by Grade-Band and Demographics—Science[8]

	Elementary School (ES)					Middle School (MS)					High School (HS)				
	N	Scale Score		Ach. Level		N	Scale Score		Ach. Level		N	Scale Score		Ach. Level	
		Mean	SD	< 3	>= 3		Mean	SD	< 3	>= 3		Mean	SD	< 3	>= 3
Student's Age															
8	260	500	56	48	212	.	.	.	.	.	.	.	.	.	.
9	464	508	69	94	370	.	.	.	.	.	.	.	.	.	.
10	225	515	75	37	188	1	–	–	–	–	.	.	.	.	.
11	.	.	.	.	.	224	516	62	64	160	.	.	.	.	.
12	.	.	.	.	.	398	513	83	132	266	1	–	–	–	–
13	.	.	.	.	.	186	513	89	66	120	.	.	.	.	.
14	.	.	.	.	.	.	.	.	.	.	3	–	–	–	–
15	.	.	.	.	.	.	.	.	.	.	344	504	85	156	188
16	.	.	.	.	.	.	.	.	.	.	8	–	–	–	–
Student's Ethnicity															
African American	450	513	64	78	372	427	517	78	137	290	190	504	91	82	108
African American/American Indian	6	–	–	–	–	4	–	–	–	–	.	.	.	.	.
American Indian	4	–	–	–	–	1	–	–	–	–	.	.	.	.	.
Asian	10	511	42	1	9	4	–	–	–	–	4	–	–	–	–
Hawaiian/Pacific Islander	1	–	–	–	–	.	.	.	.	.	.	.	.	.	.
Hispanic	40	495	66	8	32	33	493	84	10	23	10	500	22	7	3
White	414	504	70	83	331	327	513	78	107	220	149	503	80	71	78
White/African American	15	507	69	3	12	10	520	125	4	6	3	–	–	–	–
White/American Indian	1	–	–	–	–	.	.	.	.	.	.	.	.	.	.
White/Asian	3	–	–	–	–	1	–	–	–	–	.	.	.	.	.
Other	5	–	–	–	–	2	–	–	–	–	.	.	.	.	.
Unknown	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Student's Gender															
Female	311	495	66	67	244	258	512	82	82	176	122	511	80	50	72
Male	638	513	66	112	526	551	515	78	180	371	234	500	87	113	121
Unknown	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ESL (Language)															
Pre-functional	31	491	61	7	24	22	493	73	5	17	7	–	–	–	–
Beginner	1	–	–	–	–	1	–	–	–	–	.	.	.	.	.
Intermediate	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.

[8] Note: Data marked '–' are suppressed because the subgroup contains fewer than 10 students.

	Elementary School (ES)					Middle School (MS)					High School (HS)				
	N	Scale Score		Ach. Level		N	Scale Score		Ach. Level		N	Scale Score		Ach. Level	
		Mean	SD	< 3	>= 3		Mean	SD	< 3	>= 3		Mean	SD	< 3	>= 3
Advanced	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Full English Proficient	1	–	–	–	–	1	–	–	–	–	1	–	–	–	–
Title III First Year Exited	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Title III Second+ Year Exited	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
English Speaker I	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
English Speaker II	899	507	67	171	728	775	514	79	256	519	347	503	85	156	191
Pre-functional Waiver	.	.	.	.	.	1	–	–	–	–	.	.	.	.	.
Beginner Waiver	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Intermediate Waiver	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Advanced Waiver	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Unknown	17	545	65	1	16	9	–	–	–	–	1	–	–	–	–
Eligible for Free or Reduced-Price Lunch															
Free Meals	615	515	68	99	516	482	524	79	141	341	212	507	85	95	117
Full-Pay Meals	271	488	62	72	199	252	499	75	95	157	112	492	76	56	56
Reduced Meals	63	521	57	8	55	75	501	82	26	49	31	510	100	12	19
Unknown	.	.	.	.	.	.	.	.	.	.	1	–	–	–	–
IEP Disability Codes (Multiple Codes per Student)															
Severe Mental Disability	104	415	73	79	25	86	403	80	79	7	38	384	85	34	4
Moderate Mental Disability	219	500	47	35	184	274	508	50	82	192	162	512	60	73	89
Mild Mental Disability	277	542	45	5	272	196	573	58	11	185	64	571	68	7	57
Autism	217	505	58	43	174	170	496	60	71	99	55	480	62	36	19
Deaf/Blindness	8	–	–	–	–	6	–	–	–	–	5	–	–	–	–
Emotional Disability	8	–	–	–	–	3	–	–	–	–	2	–	–	–	–
Hearing Impaired	19	479	58	3	16	11	541	70	3	8	5	–	–	–	–
Learning Disability	14	583	56	.	14	14	611	94	2	12	1	–	–	–	–
Multiple-Disability	69	464	76	25	44	50	466	81	34	16	17	457	79	13	4
Other Health Impaired	57	515	65	7	50	35	527	85	10	25	19	484	79	10	9
Orthopedically Impaired	59	490	57	16	43	31	514	67	12	19	13	483	113	6	7

	Elementary School (ES)					Middle School (MS)					High School (HS)				
	N	Scale Score		Ach. Level		N	Scale Score		Ach. Level		N	Scale Score		Ach. Level	
		Mean	SD	< 3	>= 3		Mean	SD	< 3	>= 3		Mean	SD	< 3	>= 3
Speech or Language Impaired	702	513	56	104	598	407	513	69	134	273	124	511	72	59	65
Traumatic Brain Injury	4	–	–	–	–	3	–	–	–	–	1	–	–	–	–
Visually Impaired	29	442	105	19	10	31	484	78	16	15	14	429	95	10	4
TOTAL	**949**	**507**	**67**	**179**	**770**	**809**	**514**	**79**	**262**	**547**	**356**	**503**	**85**	**163**	**193**

Exhibit I-4: Performance by Grade-Band Form and Student Age—Social Studies[9]

	Elementary School					Middle School				
	N	Scale Score		Ach. Level		N	Scale Score		Ach. Level	
		Mean	SD	< 3	>= 3		Mean	SD	< 3	>= 3
Student's Age										
8	238	494	74	95	143	.	.	.	.	.
9	467	503	77	180	286	.	.	.	.	.
10	243	514	79	77	166	2	–	–	–	–
11	3	–	–	–	–	228	509	85	108	119
12	.	.	.	.	.	395	526	90	148	246
13	1	–	–	–	–	195	522	97	75	120
14										
15										
16										
Student's Ethnicity	.	.	.	.	.	.	.	.	.	.
African American	472	509	78	154	318	426	522	94	172	253
African American/American Indian	5	–	–	–	–	3	–	–	–	–
American Indian	3	–	–	–	–	3	–	–	–	–
Asian	6	–	–	–	–	11	505	77	4	7
Hawaiian/Pacific Islander	.	.	.	.	.	.	.	.	.	.
Hispanic	43	488	77	21	22	29	493	78	17	12
White	396	500	75	163	232	339	519	87	134	204
White/African American	15	502	81	5	10	5	–	–	–	–
White/American Indian	.	.	.	.	.	1	–	–	–	–
White/Asian	4	–	–	–	–	1	–	–	–	–
Other	8	–	–	–	–	2	–	–	–	–
Unknown	.	.	.	.	.	.	.	.	.	.
Student's Gender	.	.	.	.	.	.	.	.	.	.
Female	302	495	76	120	182	274	520	98	108	166
Male	650	507	77	236	413	546	520	86	225	319
Unknown	.	.	.	.	.	.	.	.	.	.
ESL (Language)										
Pre-functional	33	484	84	16	17	21	495	81	10	11
Beginner	.	.	.	.	.	.	.	.	.	.
Intermediate	.	.	.	.	.	2	–	–	–	–
Advanced	.	.	.	.	.	.	.	.	.	.
Full English Proficient	.	.	.	.	.	1	–	–	–	–
Title III First Year Exited	.	.	.	.	.	.	.	.	.	.

[9] Note: Data marked '–' are suppressed because the subgroup contains fewer than 10 students.

	Elementary School					Middle School				
	N	Scale Score		Ach. Level		N	Scale Score		Ach. Level	
		Mean	SD	< 3	>= 3		Mean	SD	< 3	>= 3
Title III Second+ Year Exited	.	.	.	.	.	.	.	.	.	.
English Speaker I	.	.	.	.	.	.	.	.	.	.
English Speaker II	901	504	77	337	563	788	520	91	320	467
Pre-functional Waiver	2	–	–	–	–	1	–	–	–	–
Beginner Waiver	.	.	.	.	.	.	.	.	.	.
Intermediate Waiver	.	.	.	.	.	.	.	.	.	.
Advanced Waiver	.	.	.	.	.	.	.	.	.	.
Unknown	16	528	52	2	14	7	–	–	–	–
Eligible for Free or Reduced-Price Lunch										
Free Meals	607	514	78	186	420	508	528	94	183	323
Full-Pay Meals	258	478	74	139	119	243	507	85	114	129
Reduced Meals	87	505	59	31	56	69	503	79	36	33
Unknown	.	.	.	.	.	.	.	.	.	.
IEP Disability Codes (Multiple Codes per Student)										
Severe Mental Disability	106	390	78	98	8	96	401	86	91	5
Moderate Mental Disability	237	496	56	101	136	277	513	55	117	160
Mild Mental Disability	249	551	50	25	224	194	595	70	11	182
Autism	226	500	63	98	127	158	504	65	84	74
Deaf/Blindness	8	–	–	–	–	5	–	–	–	–
Emotional Disability	5	–	–	–	–	2	–	–	–	–
Hearing Impaired	18	480	47	8	10	10	548	94	4	6
Learning Disability	19	569	34	.	19	7	–	–	–	–
Multiple-Disability	72	452	84	47	25	51	448	101	42	9
Other Health Impaired	60	523	68	14	46	50	536	92	15	35
Orthopedically Impaired	57	471	84	29	28	37	505	79	17	19
Speech or Language Impaired	708	507	65	254	453	422	515	77	184	238
Traumatic Brain Injury	5	–	–	–	–	4	–	–	–	–
Visually Impaired	38	439	94	27	11	29	458	99	19	10
TOTAL	**952**	**503**	**77**	**356**	**595**	**820**	**520**	**91**	**333**	**485**