

**North Carolina *EXTEND2* End-of-Grade
Science Tests**

Technical Report

***NCEXTEND2* End-of-Grade Science Tests**

March 1, 2009

In compliance with federal law, NC Public Schools administers all state-operated educational programs, employment activities and admissions without discrimination because of race, national or ethnic origin, color, age, military service, disability, or gender, except where exemption is appropriate and allowed by law.

Inquiries or complaints should be directed to:

Robert Logan, Associate State Superintendent

The Office of Innovation and School Transformation

6301 Mail Service Center

Raleigh, NC 27699-6307

919-807-3200 (phone); 919-807-4065 (fax)

Table of Contents

Chapter One: Introduction	9
1.1 Universal Participation	9
1.2 The North Carolina Testing Program	11
1.3 Participation in the North Carolina Testing Program	12
1.4 The <i>NCEXTEND2</i> Alternate Assessment	13
Chapter Two: Test Development	14
2.1 Test Development Process for the North Carolina Testing Program	14
2.2 The Curriculum Connection	16
2.3 Test Specifications	16
2.4 Item Writing.....	17
2.5 Selecting and Training Item Writers	19
2.6 Reviewing Items for Field Testing	19
2.7 Assembling Field Test Forms	20
2.8 Sampling Procedures.....	20
2.9 Item Analysis and Selection	22
2.10 Classical Measurement Theory	22
2.11 Item Response Theory (IRT) Analysis.....	22
2.12 Differential Item Functioning Analysis	23
2.13 Criteria for Inclusion in Item Pools	24
2.14 Item Statistics.....	24
2.15 Bias Review Committee	25
2.16 Pilot Testing.....	25
2.17 Operational Test Construction.....	26

2.18 Setting the Target p-value for Operational Tests	26
2.19 Setting the Estimated Test Administration Time	27
2.20 Reviewing Assembled Operational Tests	27
Chapter Three Test Administration	29
3.1 Test Administration	29
3.2 Training for Administrators	29
3.3 Preparation for Test Administration	29
3.4 Test Security and Handling Materials	29
3.5 Student Participation	30
3.6 Testing Accommodations	30
3.7 Students with Limited English Proficiency	31
3.8 Medical Exclusions	32
3.9 Reporting Student Scores	32
3.10 Confidentiality of Student Test Scores	33
Chapter Four: Scaling and Standard Setting for the <i>NCEXTEND2</i> EOG Tests Science.....	34
4.1 Conversion of Test Scores	34
4.2 Setting the Standards for the <i>NCEXTEND2</i> Tests of Science	34
4.3 Achievement Level Descriptors	35
4.4 Achievement Level Cut Scores	36
4.5 Percentile Ranking	36
Chapter Five: Reports	37
5.1 Use of Test Score Reports Provided by the North Carolina Testing Program	37
5.2 Reporting by Student	37
5.3 Reporting by School	37

5.4 Reporting by the State	38
Chapter Six: Descriptive Statistics and Reliability	39
6.1 Means and Standard Deviations for the First Operational Administration	39
6.2 Population Demographics for the First Operational Administration	39
6.3 Scale Score Frequency Distributions	42
6.4 Reliability of the North Carolina <i>EXTEND2</i> Science Tests	44
6.5 Internal Consistency of the North Carolina <i>EXTEND2</i> Science Tests	44
6.6 Standard Error of Measurement	45
6.7 Decision Consistency and Accuracy of the <i>NCEXTEND2</i> Science Tests.....	46
Chapter Seven: Evidence of Validity	47
7.1 Evidence of Validity	47
7.2 Content Validity	47
7.3 Instructional Validity.....	48
7.4 Criterion-Related Validity	48
Chapter Eight: Quality Control	50
8.1 Quality Control Prior to Test Administration	50
8.2 Quality Control in Data Preparation and Test Administration	50
8.3 Quality Control in Data Input	51
8.4 Quality Control of Test Scores	51
8.5 Quality Control in Reporting	51
Glossary of Key Terms	52
References	55
Appendix A: Item-Writing Guidelines	57
Appendix B: Test Specifications	59

Appendix C: Scale Score Frequency Distribution Tables	81
Appendix D: Testing Code of Ethics	85
Appendix E: Internal Standard Setting Meeting Attendees.....	89
Appendix F: Achievement Level Descriptors.....	90

List of Tables

Table 1: Breakdown of the number of forms and the average number of items per form for the *NCEXTEND2* 2008 field test

Table 2: Field test sample characteristics

Table 3: Average item pool statistics for the *NCEXTEND2* Science Tests

Table 4: Select statistics for the *NCEXTEND2* Science Pilot Tests

Table 5: Comparison of p-values of item pools with p-values of assembled forms

Table 6: Number of items per test and approximate administration time by grade

Table 7: Administrative Procedures Act 16 NCAC 6D .0501 (Definitions related to Student Accountability Standards)

Table 8: *NCEXTEND2* achievement levels and corresponding scale scores

Table 9: Descriptive statistics by grade for the first operational administration of the *NCEXTEND2* Tests of Science

Table 10: Population demographics for the first administration of the *NCEXTEND2* Tests of Science

Table 11: Primary exceptionality status demographics for the first operational administration of the *NCEXTEND2* Tests of Science

Table 12: Accommodations used for the first operational administration of the *NCEXTEND2* Tests of Science

Table 13: Reliability indices for the 2008 *NCEXTEND2* Tests of Science

Table 14: Reliability indexes for 2008 operational year by gender

Table 15: Reliability indexes for 2008 operational year by ethnicity

Table 16: Classical standard error of measurement by grade

Table 17: Approximate Agreement and Kappa Coefficient

Table 18: Pearson Correlation Coefficients for the *NCEXTEND2* Science Tests (Grades 5 and 8)

Table 19: Tetrachoric Correlation Coefficients for the *NCEXTEND2* Science Tests
(Grades 5 and 8)

List of Figures

Figure 1: Flow Chart of the Test Development Process used in the Development of North Carolina tests

Figure 2: Thinking skills framework used to develop items in the North Carolina Testing Program

Figure 3: Scale score frequency distribution for the 2008 Grade 5 *NCEXTEND2* Science Test

Figure 4: Scale score frequency distribution for the 2008 Grade 8 *NCEXTEND2* Science Test

Chapter One: Introduction

The General Assembly believes that all children can learn. It is the intent of the General Assembly that the mission of the public school community is to challenge with high expectations each child to learn, to achieve, and to fulfill his or her potential (G.S. 115C-105.20a).

With that mission as its guide, the State Board of Education implemented the ABCs Accountability Program at grades K–8 effective with the 1996–1997 school year and grades 9–12 effective during the 1997–1998 school year. The purpose of the assessments developed under the ABCs Accountability Program is to test students' mastery of basic skills (reading, writing, and mathematics). The ABCs Accountability Program was developed under the *Public School Laws* mandating local participation in the program, the design of annual performance standards, and the development of student academic performance standards.

1.1 Universal Participation

The School-Based Management and Accountability Program shall be based upon an accountability, recognition, assistance, and intervention process in order to hold each school and the school's personnel accountable for improved student performance in the school (G.S. 115C-105.21c).

Schools are held accountable for student learning by public reporting of student performance results on North Carolina tests. Students' scores are compiled each year and released in a report card. Schools are then recognized for the performance of their students. Schools that consistently do not make adequate progress may receive intervention from the state.

In April 1999, the State Board of Education unanimously approved Statewide Student Accountability Standards. These standards provide four Gateway Standards for student performance at grades 3, 5, 8, and 11. Students in the 3rd, 5th, and 8th grades are required to demonstrate grade-level performance in reading, writing (5th and 8th grades only), and mathematics in order to be promoted to the next grade. The law regarding student academic performance states:

The State Board of Education shall develop a plan to create rigorous student academic performance standards for kindergarten through eighth grade and student academic standards for courses in grades 9-12. The performance standards shall align, whenever possible, with the student academic performance standards developed for the National Assessment of Educational Progress (NAEP). The plan also shall include clear and

understandable methods of reporting individual student academic performance to parents (G.S. 115C-105.40).

In 2001, the reauthorization of the Elementary and Secondary Education Act (ESEA) ushered in a new era of accountability at the Federal level as well. Popularly referred to as *No Child Left Behind* (NCLB), this law was designed to improve American education by ensuring that even the neediest students receive a sound basic education and that no child is trapped in a failing school. The cornerstones of NCLB include annual testing of all students in language and mathematics in grades 3 through 8; annual testing of all students in language and math once in high school; and annual testing of all students in science in each grade span 3-5, 6-9, and 10-12. These assessment results are to be broken out (disaggregated) by ethnic, disability, poverty, and English proficiency. The end goal of NCLB is to have all students performing at a level deemed proficient, by 2014. A major provision of the Act focuses on accountability for results.

H.R. 1 will result in the creation of assessments in each state that measure what children know and learn in reading and math in grades 3-8. Student progress and achievement will be measured according to tests that will be given to every child, every year. ...

Statewide reports will include performance data disaggregated according to race, gender, and other criteria to demonstrate not only how well students are achieving overall but also progress in closing the achievement gap between disadvantaged students and other groups of students.

From: Fact Sheet on the Major Provisions of the Conference Report to H.R. 1, the No Child Left Behind Act

1.2 The North Carolina Testing Program

The North Carolina Testing Program was designed to measure the extent to which students satisfy academic performance requirements. Tests developed by the North Carolina Department of Public Instruction's Test Development Section, when properly administered and interpreted, provide reliable and valid information that enables

- *students to know the extent to which they have mastered expected knowledge and skills and how they compare to others;*
- *parents to know if their children are acquiring the knowledge and skills needed to succeed in a highly competitive job market;*
- *teachers to know if their students have mastered grade-level knowledge and skills in the curriculum and, if not, what weaknesses need to be addressed;*
- *community leaders and lawmakers to know if students in North Carolina schools are improving their performance over time and how our students compare with students from other states or the nation; and*
- *citizens to assess the performance of the public schools (North Carolina Testing Code of Ethics, 1997, revised 2000).*

The North Carolina Testing Program was initiated in response to legislation passed by the North Carolina General Assembly. The following selection from *Public School Laws* (1994) describes the legislation. *Public School Law 115C-174.10* states the following purposes of the North Carolina Testing Program:

(i) to assure that all high school graduates possess those minimum skills and that knowledge thought necessary to function as a member of society; (ii) to provide a means of identifying strengths and weaknesses in the education process in order to improve instructional delivery; and (iii) to establish additional means for making the education system at the State, local, and school levels accountable to the public for results.

Tests included in the North Carolina Testing Program are designed for use as federal, state, and local indicators of student performance. Interpretation of test scores in the North Carolina Testing Program provides information about a student's performance on the test in percentiles, scale scores, and achievement levels. Percentiles provide an indicator of how a child performs relative to other children who took the test in the norming year, or the first year the test was administered. Percentiles range from 1 to 99. A percentile rank of 65 indicates that a child performed equal to or better than 65 percent of the children who took the test during the norming year.

Scale scores are derived from a raw score or "number right" score for the test. Each test has a translation table that provides a scale score for each raw test score. Scale scores are reported alongside four achievement levels, which are predetermined academic achievement standards.

The North Carolina Testing Program, effective with the 2000-2001 school year in the Reauthorization of Individuals with Disabilities Education Act (IDEA), expanded to include a system of alternate assessments for students with disabilities.

Effective with the 2005-06 school year, the state implemented several changes in the North Carolina Testing Program as a result of changes in regulations and impending decisions by the United States Department of Education (USED). The changes included:

- 1) “Out of level” assessments are no longer permitted for any students.
- 2) Students with significant cognitive disabilities are still permitted to be assessed using an alternate assessment with alternate academic achievement standards.
- 3) The North Carolina Checklist of Academic Standards (NCCLAS) was implemented as a grade-level alternated assessment for students with disabilities and students with limited English proficiency.
- 4) The *NCEXTEND2* Alternate Assessment for EOG in reading and mathematics grades 3-8, science grades 5 and 8, and writing grades 4 and 7 was implemented for eligible students with disabilities.

1.3 Participation in the North Carolina Testing Program

All eligible students are required to participate in the North Carolina statewide testing program.

Effective in the 2005-06 school year, the North Carolina Testing Program revised its system of alternate assessments. Students in the statewide testing program have the following options for assessment.

There are three ways student may participate in the North Carolina statewide testing program:

- *Standard test administration;*
- *Standard test administration with accommodations:* On a case-by-case basis and where appropriate documentation exists, students identified as limited English proficient (LEP) and students with disabilities may receive testing accommodations. An approved list of accommodations that do not invalidate test results is provided to schools for each tested content area. The need for accommodations must be documented in the student’s appropriate LEP documentation, current IEP, or Section 504 Plan. The accommodations must be used routinely during the student’s instructional program and similar classroom assessments.
- *Alternate Assessments:* Eligible students who are limited English proficient or eligible students with disabilities who meet specific eligibility criteria can be assessed using alternate assessments.

1.4 The *NCEXTEND2* Alternate Assessment

The *NCEXTEND2* Alternate Assessment for EOG Science Grades 5 and 8 is an alternate assessment for students with disabilities who are working toward grade level achievement, but are having continued difficulty in making progress in the same time frame as students without disabilities. The assessment is a multiple-choice test with fewer items that utilizes universal design principles to address accessibility for students with disabilities. *NCEXTEND2* uses shorter reading selections, simplified language, and fewer test items and item responses (foils/answer choices) to assess students on grade-level content. *NCEXTEND2* provides access to the statewide testing through a test design that utilizes a different format and permits the use of modified academic achievement standards (achievement levels). *NCEXTEND2* Alternate Assessment for EOG Science tests are administered to students whose IEP designates *NCEXTEND2* as the appropriate assessment for End-of-Grade Science in grades 5 and 8.

This Technical Report for *NCEXTEND2* Alternate Assessment for EOG Science discusses alternate assessments aligned with the North Carolina Science 2004 *Standard Course of Study* (SCS). *NCEXTEND2* End-of-Grade Science Tests for grades 5 and 8 were field tested in Spring 2006 and administered operationally for the first time in Spring 2008.

The purpose of this document is to provide an overview of and technical documentation for the North Carolina *EXTEND2* Tests, 2nd Edition for the End-of-Grade (EOG) Science Tests in grades 5 and 8. Chapter One provides an overview of the North Carolina *EXTEND2* Tests. Chapter Two describes the test development process. Chapter Three outlines the test administration. Chapter Four describes the scoring of the tests and the standard setting process. Chapter Five provides an outline of reporting of test results. Chapters Six and Seven provide the technical properties of the tests such as descriptive statistics from the second edition, reliability indices, and evidence of validity. Chapter Eight is an overview of quality control procedures.

Chapter Two: Test Development

2.1 Test Development Process for the North Carolina Testing Program

In June of 2003, the State Board of Education codified the process used in developing all multiple-choice tests in the North Carolina Testing Program. The development of tests for the North Carolina Testing Program follows a prescribed sequence of events. A flowchart of those events is found in Figure 1.

Figure 1: Flowchart of the Test Development Process used in Development of North Carolina Tests

Curriculum Adoption	Step 7 Review Item Tryout Statistics	Step 14^b Conduct Bias Reviews
Step 1^a Develop Test Specifications (Blueprint)	Step 8^b Develop New Items	Step 15 Assemble Equivalent and Parallel Forms
Step 2^b Develop Test Items	Step 9^b Review Items for Field Test	Step 16^b Review Assembled Test
Step 3^b Review Items for Tryouts	Step 10 Assemble Field Test Forms	Step 17 Final Review of Test
Step 4 Assemble Item Tryout Forms	Step 11^b Review Field Test Forms	Step 18^{ab} Administer Test as Pilot
Step 5^b Review Item Tryout Forms	Step 12^b Administer Field Test	Step 19 Score Test
Step 6^b Administer Item Tryouts	Step 13 Review Field Test Statistics	Step 20^{ab} Establish Standards
		Step 21^b Administer Test as Fully Operational
		Step 22 Report Test Results

^aActivities done only at implementation of new curriculum

^bActivities involving NC teachers

Phase 1 (step 1) requires 4 months

Phase 2 (steps 2-7) requires 12 months

Phase 3 (steps 8-14) requires 20 months

Phase 4 (steps 15-20) requires 4 months for EOC and 9 months for EOG

Phase 5 (step 21) requires 4 months

Phase 6 (step 22) requires 1 month

TOTAL 44-49 months

NOTES: Whenever possible, item tryouts should precede field-testing items. Professional development opportunities are integral and ongoing to the curriculum and test development process.

Because of the shortened test implication schedule, the test development process for the **NCEXTEND2** End-of-Grade tests was correspondingly compressed. In this case, many of the items were items used from the general End-of-Grade tests of Science and were modified to the item format used for **NCEXTEND2**. For example, the foil deemed by content specialists to be the least plausible was removed and language was simplified when appropriate. These items had previously been through Steps 1 and 2 of the process delineated above, and in most cases, through Steps 1-14 but not used on an operational test form. The items may not have been previously used on an operational form due to lack of fit to the test specifications (i.e., too many items available in the item pool for a particular goal or objective) or target p-value of operational form. In some cases, the item had been deemed statistically “too easy,” that is a p-value greater than .90, for the general test.

2.2 The Curriculum Connection

Testing of North Carolina students’ skills relative to the competency goals and objectives in the *Standard Course of Study (SCS)* is one component of the North Carolina Testing Program. Students using the **NCEXTEND2** are tested in Science at the end of grades five and eight.

Each item on the **NCEXTEND2** End-of-Grade Science Tests is aligned to an objective from the NC *SCS* for Science for grades 5 or 8. While some objectives can be measured readily by multiple-choice questions and are assessed by the tests, other objectives address the skills and background knowledge that are needed to do well on the tests, but are not easily measured in a multiple-choice format.

2.3 Test Specifications

Delineating the purpose of a test must come before the test design. A clear statement of purpose provides the overall framework for test specifications, test blueprint, item development, tryout, and review. A clear statement of test purpose also contributes significantly to appropriate test use in practical contexts (Millman & Greene, 1993). The tests in the North Carolina Testing Program are designed in alignment with the NCSCS. The purpose of the North Carolina EOG Tests of Science is legislated by General Statute *115C-174.10* and focuses on the measurement of individual student skills and knowledge as outlined in the NCSCS.

Test specifications for the **NCEXTEND2** Science tests are developed in accordance with the competency goals and objectives specified in the NCSCS. A summary of the test specifications is provided in Appendix B. These test specifications also are generally designed to include the following:

- (1) Percentage of questions from higher or lower thinking skills and classification of each test question by level of difficulty;

- (2) Percentage of item types such as graphs, charts, diagrams, political cartoons, analogies, and other specialized constraints;
- (3) Percentage of test questions that measure a specific goal, objective, domain, or category.

In general, the *NCEXTEND2* End-of-Grade tests, were designed to follow the test specifications of the general End-of-Grade tests. Slight modifications were made to the test blueprint to accommodate the smaller number of items. The proportion of items aligned to higher versus lower order thinking skill was also modified.

Test blueprints, specific layouts or “road maps” to ensure the parallel construction of multiple test forms when necessary, were developed from the test specifications. These blueprints identify the exact numbers of items from each objective that are used in the creation of the test forms. At the objective level, the tests are comprised of items that are a random domain sample from the superordinate goal, and as such there may be more than one layout. However, at the goal level and in terms of the relative emphasis of the objective coverage, all test blueprints conform to the test specifications.

Format changes to *NCEXTEND2* from the format used for the general North Carolina EOG tests were also decided at this point. Based on observations of student work supporting the previous alternative assessment, NCAAAI, a multiple-choice format was decided upon. To reduce complexity, a two item per page, simplified language format was decided on. However, language aligned to the curriculum was retained. That is, any word in the curriculum was “fair game” for use on the assessment regardless of level of complexity. Reading load was also reduced by using three-option multiple-choice questions rather than the four-options used on the general EOG based on research supporting three options as sufficient (Haladyna & Downing, 1993).

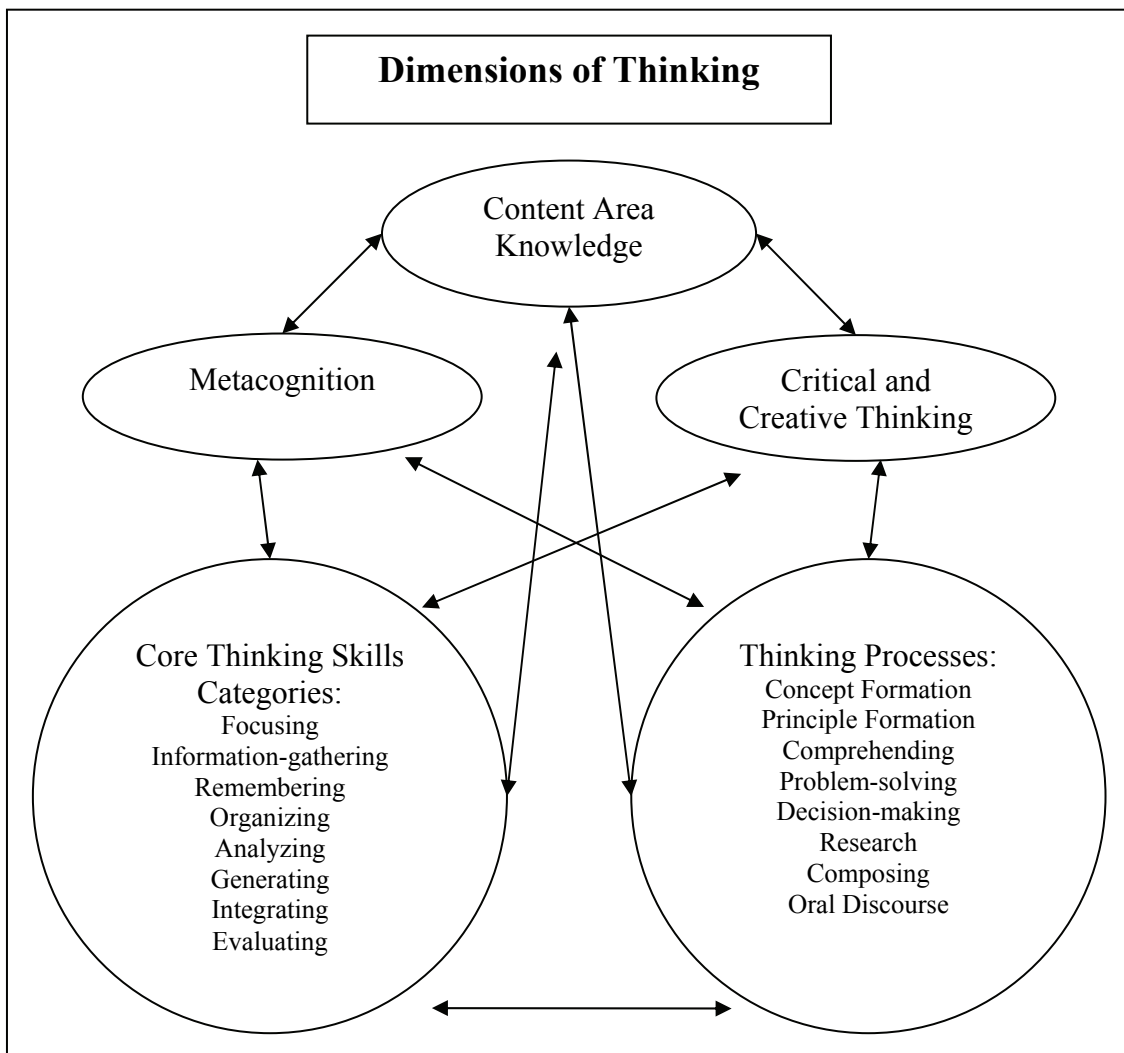
2.4 Item Writing

Each item is written to be aligned with a specific objective in the NCSCS. Items on the *NCEXTEND2* EOG Tests of Science are developed using level of difficulty and thinking skill level. Item writers use these frameworks when developing items.

For the purposes of guiding item writers to provide a variety of items, item writers were instructed to classify the items into three levels of difficulty: easy, medium, and hard. Item writers were asked to make these a priori judgments based on the population expected to be administered an *NCEXTEND2* tests as defined by the eligibility criteria. Easy items are those items that the item writer believes can be answered correctly by approximately 70% or greater of the examinees. Medium items can be answered correctly by 50–60% of the examinees. Difficult items can be answered correctly by approximately 30% or less of the examinees. The item writers were further instructed to write approximately 25% of their items at the hard level, 25% at the easy level, and the remaining 50% at the medium level of difficulty. These targets are used for item pool development to ensure an adequate range of difficulty.

A more recent consideration for item development is the classification of items by thinking skill level, the cognitive skills that an examinee must use to solve a problem or answer a test question. Thinking skill levels are based on an adaptation of *Dimensions of Thinking* by Marzano, et al. (1988). Thinking skill levels, in addition to their usefulness in framing achievement tests, also provide a practical framework for curriculum development, instruction, assessment, and staff development. Thinking skills begin with the basic skill of remembering and move to more complex thinking skills, such as analysis, integration, and evaluation. Figure 2 below shows a visual representation of the framework.

Figure 2: Thinking skills framework used to develop items in the North Carolina Testing Program adapted from Robert Marzano (1988).



2.5 Selecting and Training Item Writers

Once the test specifications were outlined for the *NCEXTEND2* EOG Tests of Science, North Carolina educators were recruited and trained to write new items for the state tests. As addressed in section 2.1, in some cases for the EOG *NCEXTEND2*, items were modified by internal content specialists from previously written items for the general EOG.

Diversity among the item writers and their knowledge of the current NCSCS was addressed during recruitment. The purpose of using North Carolina educators to develop items was to ensure instructional validity of the items. For some items for the *NCEXTEND2* End-of Grade tests, some item development was contracted to an external vendor. The vendor was encouraged to use North Carolina educators in addition to professional item writers to generate items that would align with the NCSCS.

Item writers received a packet of materials designed from the Science curriculum, which included information on content and procedural guidelines as well as information on stem and foil development. The item writing guidelines are included in Appendix A. The items developed during the training were evaluated by content specialists, who then provided feedback to the item writers on the quality of their items.

2.6 Reviewing Items for Field Testing

Each item was reviewed by North Carolina educators prior to being placed on a field test. Once items were reviewed by educators, test development staff members, with input from curriculum specialists, reviewed each item. Items were also reviewed by educators and/or staff members who are familiar with the needs of students with disabilities and students with limited English proficiency.

The criteria used by the review team to evaluate each test item included the following:

- 1) Conceptual criteria:
 - objective match (curricular appropriateness)
 - thinking skill match
 - fair representation
 - lack of bias
 - clear statement
 - single problem
 - one best answer
 - common context in foils
 - each foil credible
 - meets all technical criteria for item parameters
- 2) Language criteria:
 - appropriate for age, simplified as appropriate

- correct punctuation
- spelling and grammar
- lack of excess words
- no stem/foil clues
- no negative in foils

3) Format criteria:

- logical order of foils
- familiar presentation style, print size, and type
- correct mechanics and appearance
- equal length foils

4) Diagram criteria:

- necessary
- clean
- relevant
- unbiased

The detailed review of items prior to field testing helped to prevent the loss of items due to quality issues.

2.7 Assembling Field Test Forms

When developing tests for the North Carolina Testing Program, items written for each grade and subject area were assembled into forms for field testing. The forms were organized according to specifications set forth for the operational tests. Additional teachers reviewed the assembled forms for clarity, correctness, potential bias, and curricular appropriateness. The following table provides a breakdown of the number of forms and the number of items per form for the Spring 2006 field test.

Table 1: Breakdown of the number of forms and the number of average items per form for the *NCEXTEND2* 2006 field test

Subject/ Grade	Number of Forms	Number of Items per Form
Science 5	3	40
Science 8	3	40

2.8 Sampling Procedures

Items for the test were field tested using the population of students for which the assessment was intended. The following table provides a breakdown of the field test population.

Table 2: Field test sample characteristics

Course/ Grade	Year	Number of students tested	Gender		Ethnicity					
			% Female	% Male	% Asian	% Black	% Hispanic	% American Indian	% Multi	% White
Science 5	2006	2129	31.81	68.19	0.57	46.41	9.78	3.21	2.27	37.76
Science 8	2006	1658	34.26	65.74	0.79	47.61	6.89	2.60	2.18	39.88

Notes: The percentages for demographic categories are for all examinees with available demographic data, averaged across forms.

2.9 Item Analysis and Selection

Field testing provides important data for determining whether an item will be retained for use on an operational *NCEXTEND2* EOG Test of Science. The North Carolina Testing Program typically uses both classical measurement analysis and item response theory analysis to determine if an item has sound psychometric properties. These analyses provide information that assists North Carolina Testing Program staff and consultants in determining the extent to which an item can accurately measure a student's level of achievement.

Field-test data for the *NCEXTEND2* tests were analyzed by the NCDPI psychometric staff. Item statistics and description information were then attached to the item record for each item. Item records contained the statistical, descriptive, and historical information for an item; a copy of the item as it was field-tested; any comments by reviewers; and curricular and psychometric notations.

2.10 Classical Measurement Analysis

For each item, the p-value (proportion of examinees answering an item correctly) and the point-biserial correlation between the item score and the total test score were computed using the statistical software SAS. In addition, frequency distributions of the response choices were tabulated.

2.11 Item Response Theory (IRT) Analysis

To provide additional information about item performance, the North Carolina Testing Program also uses IRT statistics to determine whether an item should be included on the test. IRT is, with increasing frequency, being used with large-scale achievement testing. "The reason for this may be the desire for item statistics to be independent of a particular group and for scores describing examinee proficiency to be independent of test difficulty, and for the need to assess reliability of tests without the tests being strictly parallel" (Hambleton, 1983, p. 148). Regardless of the distribution of the sample, the parameter estimates will be linearly related to the parameters estimated with some other sample drawn from the same population. IRT allows the comparison of two students' ability estimates even though they may have taken different items. An important characteristic of IRT is item-level orientation. IRT makes a statement about the relationship between the probability of answering an item correctly and the student's ability or the student's level of achievement. The relationship between a student's item performance and the set of traits underlying item performance can be described by a monotonically increasing function called an Item Characteristic Curve (ICC). This function specifies that as the level of the trait increases, the probability of a correct response to an item increases.

Due to the smaller sample size of the *NCEXTEND2* population, the one-parameter logistic model (1PL) of IRT was used in generating the IRT-based difficulty. This model

takes into account the difficulty of the item and the ability of the examinee. A student's probability of answering a given item correctly depends on the student's ability and the characteristics of the item. The 1PL model has three assumptions:

- (1) unidimensionality—only one ability is assessed by the set of items (for example, a spelling test only assesses a student's ability to spell);
- (2) local independence—when abilities influencing test performance are held constant, an examinee's responses to any pair of items are statistically independent (conditional independence, i.e., the only reason an examinee scores similarly on several items is because of his or her ability); and
- (3) the ICC specified reflects the true relationship among the unobservable variable (ability) and the observable variable (item response).

The formula for the 1PL model is

$$P_i(\theta) = \frac{e^{Da(\theta-b_i)}}{1 + e^{Da(\theta-b_i)}}$$

where

- $P_i(\theta)$ —the probability that a randomly chosen examinee with ability (θ) answers item i correctly (this is an S-shaped curve with values between 0 and 1 over the ability scale)
- a —the slope or the discrimination power of the item (a constant)
- b —the threshold, or “difficulty parameter,” the point on the ability scale where the probability of a correct response is 50%
- D —a scaling factor, 1.7, to make the logistic function as close as possible to the normal ogive function (Hambleton, 1983, p.125).

The IRT parameter estimates for each item are computed using the BILOG computer program (Muraki, Mislevy, & Bock, 1991) using the default Bayesian prior distributions for the item parameters.

2.12 Differential Item Functioning Analysis

It is important to know the extent to which an item on a test performs differently for different students. Differential item functioning (DIF) examines the relationship between the score on an item and group membership while controlling for ability. The Mantel-Haenszel procedure examines DIF by examining ($j \times 2 \times 2$) contingency tables, where j is the number of different levels of ability actually achieved by the examinees (actual total scores received on the test). The focal group is the focus of interest and the reference group serves as a basis for comparison for the focal group (Dorans and Holland, 1993; Camilli and Shepherd, 1994). For example, females might serve as the focal group and

males might serve as the reference group to determine if an item is biased toward or against females.

The Mantel-Haenszel chi-square statistic tests the alternative hypothesis that a linear association exists between the row variable (score on the item) and the column variable (group membership). The χ^2 distribution has one degree of freedom (*df*) and is determined where r^2 the Pearson correlation between the row variable and the column variable (*SAS Institute, 1994*). The Mantel-Haenszel (MH) Log Odds Ratio statistic in SAS was used to determine the direction of differential item functioning (DIF). This measure was obtained by combining the odds ratios across levels with the formula for weighted averages (Camilli and Shepherd, 1994, p. 110).

For this statistic, the null hypothesis of no relationship between score and group membership, or that the odds of getting the item correct are equal for the two groups, is not rejected when the odds ratio equals 1. For odds ratios greater than 1, the interpretation is that an individual at score level *j* of the reference group has a greater chance of answering the item correctly than an individual at score level *j* of the focal group. Conversely, for odds ratios less than 1, the interpretation is that an individual at score level *j* of the focal group has a greater chance of answering the item correctly than an individual at score level *j* of the reference group. The Breslow-Day Test is used to test whether the odds ratios from the *j* levels of the score are all equal. When the null hypothesis is true, the statistic is distributed approximately as a chi square with *j*-1 degrees of freedom (*SAS Institute, 1985*).

2.13 Criteria for Inclusion in Item Pools

Items were flagged as exhibiting psychometric problems or differential item functioning due to ethnicity/race or gender according to the following criteria:

- “weak prediction”—the *p*-bis was less than .25
- ethnic DIF—the log odds ratio was greater than 1.5 (favored whites) or less than 0.67 (favored blacks), and
- gender DIF—the log odds ratio was greater than 1.5 (favored females) or less than 0.67 (favored males).

Because the tests were to be used to evaluate the implementation of the curriculum, items were not flagged on the basis of the difficulty of the item (threshold).

2.14 Item Statistics

Final average item pool characteristics for each of the *NCEXTEND2* Tests are provided in the following chart.

Table 3: Average item pool statistics for the *NCEXTEND2* Science Tests.

Subject/ Grade	p-value	r-bis	Bias (Odds Ratio)	
			<i>Gender</i>	<i>Ethnic/Race</i>
Science 5	0.417	0.130	1.009	1.045
Science 8	0.447	0.172	1.028	1.068

All items, statistics, and comments were reviewed by curriculum specialists and testing consultants, and items found to be inappropriate for curricular or psychometric reasons were deleted. In addition, items flagged for exhibiting ethnic or gender DIF (Mantel-Haenszel indices greater than 1.5 or less than 0.67) were then reviewed by a bias review committee as described below.

2.15 Bias Review Committee

The bias review team members, selected because of their knowledge of the curriculum area and their diversity, evaluated the items using the following questions:

- 1) Does the item contain any offensive gender, ethnic, religious, or regional content?
- 2) Does the item contain gender, ethnic, or cultural stereotyping?
- 3) Does the item contain activities that will be more familiar to one group than another?
- 4) Do the words in the item have a different meaning in one group than in another?
- 5) Could there be group differences in performance that are unrelated to proficiency in the content areas?

An answer of “yes” to any of the questions resulted in the unique five-digit item number being recorded on an item bias sheet along with the nature of the bias.

Items that were flagged by the bias review committee were then reviewed by curriculum specialists. If curriculum found the items measured content expected to be mastered by all students, the item was retained for test development. Items consistently identified as exhibiting bias by both review committees were deleted from the item pool.

2.16 Pilot Testing

Following field testing of items, pilot tests were constructed. For all *NCEXTEND2* Science tests one pilot form was assembled from items that were found to be psychometrically sound and measure curriculum standards as specified in the test specifications. The forms for each grade were developed according to test specifications outlined during the initial phase of test development.

Following administration, the 48-item pilot tests were reviewed for both test- and item-level statistical performance. Due to concerns about the reliability of the assessment, scores were not released and the number of items to go on the operational test was increased from 48 to 60.

Table 4: Select statistics for the *NCEXTEND2* Science Pilot Tests

Subject/ Grade	Number of items	p-Value of Assembled Pilot Form	average p-value after administration	Coefficient alpha
Science 5	48	0.474	.510	0.6927
Science 8	48	0.474	.518	0.7903

2.17 Operational Test Construction

Following field testing of items, operational tests were constructed. For all *NCEXTEND2* Science tests one operational form was assembled from items that were found to be psychometrically sound and measure curriculum standards as specified in the test specifications. The final item pool was based on approval by the (1) TOPS Content experts for curriculum purposes and (2) NCDPI Division of Accountability Services/NC Testing Program for psychometrically sound item performance. The forms for each grade and course were developed according to test specifications outlined during the initial phase of test development.

2.18 Setting the Target p-value for Operational Tests

P-value is a measure of the difficulty of an item. P-values can range from 0 to 1. The letter “p” symbolizes the proportion of examinees that endorse an item correctly. So an item with a p-value of 0.75 was correctly endorsed by 75 percent of the students who took the item during the field test, and one might expect that roughly 75 of the 100 will examinees answer it correctly when the item is put on an operational test. An easy item has a p-value that is high—that means that a large proportion of the examinees got the item right during the field test. A difficult item has a low p-value, meaning that few examinees endorsed the item correctly during tryout.

The NCDPI psychometric staff must choose a target p-value for each operational test prior to assembling the tests. Ideally, the average p-value of a test of a 3-option test would be 0.665, which is the theoretical average of a student getting 100 percent correct on the test and a student scoring a “chance” performance (33 percent for a 3-option multiple-choice test). That is, $(100 + 33/2)$. The target is chosen by first looking at the distribution of the p-values for a particular item pool. While the goal is to set the target as close to 0.665 as possible, it is often the case that the target p-value is set between the ideal 0.665 and the average p-value of the item pool. The average p-value of the item pool and the p-value of assembled forms are provided for comparison.

Table 5: Comparison of p-values of item pools with p-values of assembled forms

Subject/ Grade	p-Value of Item Pool*	p-Value of Assembled Pilot Form*	p-Value after pilot administration
Science 5	0.417	0.462	0.497
Science 8	0.447	0.486	0.538

*statistics from Spring 06 field test

2.19 Setting the Estimated Test Administration Time

Other important considerations in the construction of the *NCEXTEND2* tests were the number of items to be included on the test and the time necessary to complete testing. When assembling operational tests, the NCDPI psychometric staff reviewed field test timing data. They determined the amount of time it should take the majority of the students to complete the test. For operational tests, the total number of items and the approximate administration time for each grade/subject area is provided below.

Table 6: Number of items per test and approximate administration time by grade

Subject/ Grade	Number of Items	Approximate Time
Science 5	60	150 minutes
Science 8	60	150 minutes

However, it is noted that students are required to receive ample opportunity to complete the science tests. As long as the students are engaged and working, they must be allowed ample time to complete the test.

2.20 Reviewing Assembled Operational Tests

Once forms were assembled to meet test specifications, target p-values, and item parameter targets, TOPS content staff, TOPS EC staff, subject area teachers and EC teachers then reviewed the assembled forms. The criteria for evaluating each group of forms included the following:

- ❑ The content of the test forms should reflect the goals and objectives of the North Carolina *Standard Course of Study* for the subject and grade (curricular validity);
- ❑ The content of test forms should reflect the goals and objectives taught in North Carolina schools (instructional validity);

- Items should be clearly and concisely written, and the vocabulary appropriate to the target age level (item quality);
- Content of the test forms should be balanced in relation to ethnicity, gender, socioeconomic status, and geographic district of the state (test/item bias); and
- Each item should have one and only one best answer that is right; however, the distractors should appear plausible for someone who has not achieved mastery of the representative objective (one best answer).

Reviewers were instructed to take the tests (circling the correct responses in the booklet) and to provide comments and feedback next to each item. After reviewing the form, each reviewer independently completed the survey asking for his or her opinion as to how well the tests met the five criteria listed above. During the last part of the session the group discussed the tests and made comments as a group. The ratings and the comments were aggregated for review by NCDPI curriculum specialists and testing consultants. Test development staff members, with input from curriculum staff and content experts, and editors conducted the final content and grammar check for each test form.

Chapter Three: Test Administration

3.1 Test Administration

The *NCEXTEND2* End-of-Grade Science Tests are administered to students in grades 5 and 8 who meet eligibility criteria as part of the statewide assessment program. The standard for grade-level proficiency is a test score at Achievement Level Three or above.

3.2 Training for Administrators

The North Carolina Testing Program uses a train-the-trainer model to prepare test administrators to administer North Carolina tests. Regional Accountability Coordinators (RACs) receive training in test administration from the NCDPI Testing Policy and Operations staff at regularly scheduled monthly training sessions. Subsequently, the RACs provide training on conducting a proper test administration to Local Education Agency (LEA) test coordinators. LEA test coordinators provide training to school test coordinators. The training includes information on the test administrators' responsibilities, proctors' responsibilities, preparing students for testing, eligibility for testing, policies for testing students with special needs (students with disabilities and students with limited English proficiency), test security (storing, inventorying, and returning test materials), and the *Testing Code of Ethics* (see Appendix D).

3.3 Preparation for Test Administration

School test coordinators must be accessible to test administrators and proctors during the administration of secure state tests. The school test coordinator is responsible for monitoring test administrations within the building and responding to situations that may arise during test administrations. Only employees of the school system are permitted to administer secure state tests. Test administrators are school personnel who have professional training in education and the state testing program. Test administrators may not modify, change, alter, or tamper with student responses on the answer sheets or test books. Test administrators are to thoroughly read the *Test Administrator's Manual* prior to actual test administration; discuss with students the purpose of the test; and read and study the codified North Carolina *Testing Code of Ethics*.

3.4 Test Security and Handling Materials

Compromised secure tests result in compromised test scores. To prevent contamination of test scores, the NCDPI maintains test security before, during, and after test administration at both the school system level and the individual school. School systems

are also mandated to provide a secure area for storing tests. The Administrative Procedures Act 16 NCAC 6D .0302. states, in part, that

school systems shall (1) account to the department (NCDPI) for all tests received; (2) provide a locked storage area for all tests received; (3) prohibit the reproduction of all or any part of the tests; and (4) prohibit their employees from disclosing the content of or discussing with students or others specific items contained in the tests. Secure test materials may only be stored at each individual school for a short period prior to and after the test administration. Every effort must be made to minimize school personnel access to secure state tests prior to and after each test administration.

At the individual school, the principal shall account for all test materials received. As established by APA 16 NCAC 6D .0306, the principal shall store test materials in a secure, locked area except when in use. The principal shall establish a procedure to have test materials distributed immediately prior to each test administration. Before each test administration, the building level coordinator shall collect, count, and return all test materials to the secure, locked storage area. Any discrepancies are to be reported to the school system test coordinator immediately and a report must be filed with the regional accountability coordinator.

3.5 Student Participation

The Administrative Procedures Act 16 NCAC 6D. 0301 requires that all public school students in enrolled grades for which the SBE adopts a test, including every child with disabilities, shall participate in the testing program unless excluded from testing as provided by 16 NCC 6G.0305(g).

NCEXTEND2 End-of-Grade Science Tests (Grades 5 and 8)

All students in membership in grades 5 and 8 with an IEP that designates the use of **NCEXTEND2** for testing purposes are required to participate in the **NCEXTEND2** End-of-Grade Science Tests.

3.6 Testing Accommodations

On a case-by-case basis where appropriate documentation exists, students with disabilities and students with limited English proficiency may receive testing accommodations. The need for accommodations must be documented in a current Individualized Education Program (IEP), Section 504 Plan, or LEP Plan. The accommodations must be used routinely during the student's instructional program or similar classroom assessments. For information regarding appropriate testing procedures, test administrators who provide accommodations for students with disabilities must refer

to the most recent publication of *Testing Students with Disabilities* and any published supplements or updates. The publication is available through the local school system or at www.ncpublicschools.org/accountability/testing. Test administrators must be trained in the use of the specified accommodations by the school system test coordinator or designee prior to the test administration.

3.7 Students with Limited English Proficiency

Per HSP-C-005, students identified as limited English proficient shall be included in the statewide testing program. Students identified as limited English proficient who have been assessed on the state-identified language proficiency test as below Intermediate High in reading may participate for up to 2 years (24 months) in U.S. schools in the NCCLAS as an alternate assessment in the areas of reading and mathematics at grades 3 through 8 and 10 and in high school courses in which an end-of-course test is administered. Students identified as limited English proficient who have been assessed on the state-identified language proficiency test as below Superior, per HSP-A-011, in writing may participate in the NCCLAS in writing for grades 4, 7, and 10 for up to 2 years (24 months) in U.S. schools. All students identified as limited English proficient must be assessed using the state-identified language proficiency test at initial enrollment and annually thereafter during the window of February 1 to April 30. A student who enrolls after January 1 does not have to be retested during the same school year. Limited English proficient students who are administered the NCCLAS shall not be assessed off-grade level. In March 2004, the State Board of Education adopted a temporary rule to make the following changes with respect to limited English proficient students during their first year in U.S. schools.*

**Note: First year of enrollment in U.S. schools refers to the first school year that a student has been enrolled in a U.S. school. It does not refer to a 12-month period. If a student has been enrolled in any U.S. school prior to this school year, the student, regardless of his/her enrollment period would be expected to be assessed in reading and mathematics.*

Schools shall:

- continue to administer state reading and mathematics tests for LEP students who score at or above Intermediate High on the reading section of the language proficiency test during their first year in U.S. schools. Results from these assessments will be included in the ABCs and AYP.
- not require LEP students (who score below Intermediate High on the reading section of the language proficiency test) in their first year in U.S. schools to be assessed on the reading End-of-Grade tests, High School Comprehensive Test in Reading, or the NC Checklist of Academic Skills (NCCLAS) for reading.
- for purposes of determining the 95% tested rule in reading, use the language proficiency test from the spring administration for these students.

- not count mathematics results in determining AYP or ABCs performance composite scores for LEP students who score below Intermediate High on the reading section of the language proficiency test in their first year in U.S. schools.
- include students previously identified as LEP, who have exited LEP identification during the last two years, in the calculations for determining the status of the LEP subgroup for AYP only if that subgroup already met the minimum number of 40 students required for a subgroup.

3.8 Medical Exclusions

In some rare cases, students may be excused from the required state tests. The process for requesting special exceptions based on significant medical emergencies and/or conditions is as follows:

For requests that involve significant medical emergencies and/or conditions, the LEA superintendent or charter school director is required to submit a justification statement that explains why the emergency and/or condition prevents participation in the respective test administration during the testing window and the subsequent makeup period. The request must include the name of the student, the name of the school, the LEA code, and the name of the test(s) for which the exception is being requested. Medical documents are not included in the request to NCDPI. The request is to be based on information housed at the central office. The student's records must remain confidential. Requests must be submitted prior to the end of the makeup period for the respective test(s). Requests are to be submitted for consideration by the LEA superintendent or charter.

3.9 Reporting Student Scores

According to APA 16 NCAC 6D .0302, schools systems shall, at the beginning of the school year, provide information to students and parents or guardians advising them of the districtwide and state-mandated tests that students will be required to take during the school year. In addition, school systems shall provide information to students and parents or guardians to advise them of the dates the tests will be administered and how the results from the tests will be used. Also, information provided to parents about the tests shall include whether the State Board of Education or local board of education requires the test. School systems shall report scores resulting from the administration of the districtwide and state-mandated tests to students and parents or guardians along with available score interpretation information within 30 days from the generation of the score at the school system level or receipt of the score and interpretive documentation from the NCDPI.

At the time the scores are reported for tests required for graduation such as competency tests and the computer skills tests, the school system shall provide information to students and parents or guardians to advise whether or not the student has met the standard for the test. If a student fails to meet the standard for the test, the students and parents or guardians shall be informed of the following at the time of reporting: (1) the date(s) when focused remedial instruction will be available and (2) the date of the next testing opportunity.

3.10 Confidentiality of Student Test Scores

State Board of Education policy states that “any written material containing the identifiable scores of individual students on tests taken pursuant to these rules shall not be disseminated or otherwise made available to the public by any member of the State Board of Education, any employee of the State Board of Education, the State Superintendent of Public Instruction, any employee of the North Carolina Department of Public Instruction, any member of a local board of education, any employee of a local board of education, or any other person, except as permitted under the provisions of the Family Educational Rights and Privacy Act of 1974, 20 U.S.C. § 1232g.”

Chapter Four: Scaling and Standard-Setting for the *NCEXTEND2* EOG Tests of Science

The *NCEXTEND2* Tests' scores are reported as scale scores, achievement levels, and percentiles. There are several advantages to using scale scores:

- ❑ Scale scores can be used to compare test results when there have been changes in the curriculum and/or changes in the method of testing.
- ❑ Scale scores on pretests or released test forms can be related to scale scores used on secure test forms administered at the end of the course.
- ❑ Scale scores can be used to compare the results of tests that measure the same content area but are composed of items presented in different formats.
- ❑ Scale scores can be used to minimize differences among various forms of the tests.

4.1 Conversion of Test Scores

For the *NCEXTEND2* EOG Tests of Science, each student's score is determined by calculating the number of items he or she answered correctly and then converting the sum to a scale score using the one-parameter IRT model. Items are assigned a score of 0 if the student did not answer the item correctly and a score of 1 if the student did answer the item correctly. Software developed at the L.L. Thurstone Psychometric Laboratory at the University of North Carolina at Chapel Hill converts raw scores (total number of items answered correctly) to scale scores using the three IRT parameter for each item. The software implements the algorithm described by Thissen and Orlando (2001, pp. 119-130).

4.2 Setting the Standards for the *NCEXTEND2* Tests of Science

The NCDPI has a history of utilizing multiple sources of information in regard to setting academic achievement standards, or "cut scores," for tests in the statewide testing program. Historically, the process of standard setting has involved two phases. The first phase of standard setting is where the state gathers information from teachers regarding expected performance of students. This information gathering phase includes an examinee-centered and/or test-based method of standard setting. The second phase of standard setting involves policymakers (State Board of Education) and other stakeholders within the NCDPI, such as curriculum, exceptional children, and assessment staff. Stakeholders within the NCDPI take the recommendations from the teacher panels and consolidate the information to present to the State Board of Education for final approval.

Ultimately, the setting of academic achievement standards is a policy decision that is made by the State Board of Education.

For the *NCEXTEND2* Tests of Science, standard setting or the process of determining “cut scores” for the different achievement levels was typically accomplished through the use of Modified Bookmark, or item mapping, standard setting. Modified Bookmark is a test-based method of standard setting. In this technique, an appropriate group (either an expert panel, a representative group of users, or a policymaker group) reviews a specially-constructed test booklet that is arranged in item difficulty order (Lewis, Mitzel, & Green, 1996). The standard setter (panelist) is asked to mark the spot in a specially constructed test booklet (arranged in order of item difficulty) where a set percentage of minimally-proficient or minimally-advanced students would pass the item. An alternate method is for the standard setter (panelist) to mark where the difference in performance of the proficient and advanced student on an exercise is a desired minimum percentage of students.

Internal meetings were held following the standard setting process to review the cut scores set using this method. Participants from the internal meeting are provided in Appendix E. Additional information using the contrasting groups standard setting methodology was provided to the internal meeting attendees. All meeting participants (except for staff psychometricians, who abstained) voted to recommend the cut scores lowered by two standard errors at the proficiency cut to the North Carolina State Board of Education. The SBE is the ultimate decision maker on cut scores. Cut scores were approved on October 2, 2008.

4.3 Achievement Level Descriptors

The four achievement levels in the North Carolina Student Accountability System are operationally defined below.

Table 7: Administrative Procedures Act 16 NCAC 6D .0501 (Definitions related to Student Accountability Standards)

Achievement Levels for the North Carolina Testing Program	
Level I	Students performing at this level do not have sufficient mastery of knowledge and skills in this subject area to be successful at the next grade level.
Level II	Students performing at this level demonstrate inconsistent mastery of knowledge and skills that are fundamental in this subject area and that are minimally sufficient to be successful at the next grade level.
Level III	Students performing at this level consistently demonstrate mastery of grade level subject matter and skills and are well prepared for the next grade level.
Level IV	Students performing at this level consistently perform in a superior manner clearly beyond that required to be proficient at grade level work.

During the modified Bookmark standard setting, a fuller, more descriptive Achievement Level Descriptor was created by the panelists. This document was intended to describe what the student should be able to do, in terms of knowledge of the content area, at each Achievement Level. These Achievement Level Descriptors are found in Appendix G.

4.4 Achievement Level Cut Scores

The achievement level score ranges for the *NCEXTEND2* North Carolina Science Tests are provided below.

Table 8: *NCEXTEND2* achievement levels and corresponding scale scores

Grade/Subject	Level I	Level II	Level III	Level IV
Science 5	≤144	145-156	157-164	≥165
Science 8	≤140	141-154	155-160	≥161

4.5 Percentile Ranking

The percentile rank for each scale score is the percentage of scores less than or equal to that score. If the percentile formula is applied to the frequency distribution of scores for grade 5 Science, then a score of 260 would have a percentile rank of 89. Eighty-nine percent of students scored at or below a score of 260. The percentile rank provides information about a student's score on a test relative to other students in the norming year. The percentile ranks are calculated after the first operational year.

Score reports are generated at the local level to depict achievement for individual students, classrooms, schools, and local education agencies. These data can be disaggregated by subgroups of gender and race/ethnicity as well as other demographic variables collected during the test administration. Demographic data are reported on variables such as free/reduced lunch status, LEP status, migrant status, Title I status, disability status, and parents' levels of education. The results are reported in aggregate at the state level usually at the end of June of each year. The NCDPI uses these data for school accountability and to satisfy other federal requirements such as Adequate Yearly Progress (AYP) and federal *No Child Left Behind* mandates.

Chapter Five: Reports

5.1 Use of Test Scores Reports Provided by the North Carolina Testing Program

The North Carolina Testing Program provides reports at the student level, school level, and state level. The North Carolina *Testing Code of Ethics* dictates that educators use test scores and reports appropriately. This means that educators recognize that a test score is only one piece of information and must be interpreted together with other scores and indicators. Test data help educators understand educational patterns and practices. Data analysis of test scores for decision-making purposes should be based upon disaggregation of data by student demographics and other student variables as well as an examination of grading practices in relation to test scores, growth trends, and goal summaries for state-mandated tests.

5.2 Reporting by Student

The state provides scoring equipment in each school system so that administrators can score all state-required multiple-choice tests. This scoring generally takes place within two weeks after testing so the individual score report can be given to the student and parent before the end of the school year.

Each student at grades 5 and 8 who takes the *NCEXTEND2* End-of-Grade Science test is given a “Parent/Teacher Report.” This single sheet provides information on that student’s performance on the science test. A flyer titled, “Understanding the Individual Student Report for the *NCEXTEND2* EOG Grades 3-8,” is provided with each “Parent/Teacher Report.” This publication offers information for understanding student scores as well as suggestions on what parents and teachers can do to help students in the area of science.

The student report also shows how that student’s performance compared to the average scores for the school, the school system, and the state. A four-level achievement scale is used for the tests:

Achievement Level I represents insufficient mastery of the subject.

Achievement Level II is inconsistent mastery of the subject.

Achievement Level III is consistent mastery and the minimum goal for students.

Achievement Level IV is superior mastery of the subject.

5.3 Reporting by School

Since 1997, the student performance on end-of-grade tests for each elementary and middle school has been released by the state through the ABCs School Accountability. High school student performance began to be reported in 1998 in the ABCs School

Accountability. For each school, parents and others can see the actual performance for groups of students at the school in reading, mathematics, and writing; the percentage of students tested; whether the school met or exceeded goals that were set for it; and the status designated by the state. Starting in 2008-09, the ABC's are proposed to include ***NCEXTEND2*** tests of Science in the performance composite of the ABC model. Student growth in Science will not be included in the model. The ABCs report is available on the Department of Public Instruction Web site at <http://abcs.ncpublicschools.org/abcs/>. School principals also can provide information about the ABC report to parents.

5.4 Reporting by the State

The state reports information on student performance in various ways. The North Carolina Report Cards provide information about K-12 public schools (including charters and alternative schools) for schools, school systems, and the state. Each report card includes a school or district profile and information about student performance, safe schools, access to technology, and teacher quality.

As a participating state in the National Assessment of Educational Progress (NAEP), North Carolina student performance on selected subjects is included in annual reports released nationally. The state also releases state and local SAT scores each summer.

Chapter Six: Descriptive Statistics and Reliability

6.1 Means and Standards Deviations for the First Operational Administration

The North Carolina *EXTEND2* Science Tests were administered operationally for the first time in the spring of 2008. Descriptive statistics for the first operational year of the *NCEXTEND2* EOG Tests of Science are provided below.

Table 9: Descriptive statistics by grade for the first operational administration of the *NCEXTEND2* Tests of Science

Year	Subject/Grade	N students tested	Mean	Standard Deviation
2008	Science 5	3004	149.95	8.500
2008	Science 8	2678	150.13	8.998

6.2 Population Demographics for the First Operational Administration

The population of the students who took the first operational administration of the North Carolina *EXTEND2* Science Test in Spring 2008 are described in the tables below.

Table 10: Population demographics for the first operational administration of the *NCEXTEND2* Tests of Science

Grade/Course	N students tested	Female	Male	Asian	Black	Hispanic	American Indian	White	Multi-Racial
Science 5	3004	34.19	65.81	0.83	40.89	11.73	2.57	3.43	40.55
Science 8	2678	32.25	67.75	0.9	43.62	7.9	2.56	2.37	42.64

Table 11: Primary exceptionalality status demographics for the first operational administration of the *NCEXTEND2* Tests of Science

Subject/ Grade	N	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Science 5	3004	5.23	0.03	0.03	4.20	1.47	30.49	0.80	0.00	34.82	0.23	17.53	0.63	0.57	0.40	0.13
Science 8	2678	4.44	0.04	0.00	4.40	0.90	40.50	1.05	0.00	26.19	15.92	15.92	0.38	0.26	0.26	0.15

- 1= Autism
- 2= Deaf/Blind
- 3= Deafness
- 4= Serious Emotional Disability
- 5= Hearing Impaired
- 6= Intellectual Disability- Mild
- 7= Intellectual Disability- Moderate
- 8= Intellectual Disability- Serious
- 9= Specific Learning Disabled
- 10= Multihandicapped
- 11= Other Health Impaired
- 12= Orthopedically Impaired
- 13= Speech-Language Impaired
- 14=Traumatic Brain Injured
- 15= Visually Impaired

Table 12: Accommodations used for the first operational administration of the *NCNEXTEND2* Tests of Science

Subject/Grade	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
Science 5		19	21	1		16	3	3	2746	2547	0	1168	49	2275	6	30	2819		31
Science 8	3	15	9		1	4	3	2	2447	2402	14	681	31	1315	22	18	2397		10

- 1= Braille Edition
- 2= Large Print Edition
- 3= Assistive Technology Devices
- 4= Keyboarding Devices
- 5= Crammer Abacus
- 6= Dictation to Scribe
- 7= Magnification Device
- 8= Hospital/Home Testing
- 9= Testing in a Separate Room
- 10= Scheduled Extended Time

- 11= One Test Item Per Page Edition
- 12= Multiple Testing Sessions
- 13= Student Reads Test Aloud to Self
- 14= Student Marks in Test Book
- 15= Bilingual Dictionary/Electronic Translator
- 16= Interpreter/Tranliterator Signs/Cues Test**
- 17= Test Administrator Reads Test Aloud (in English)
- 18= Braille Writer/State & Stylus
- 19= Accommodation Notification Form

**Use of the "Interpreter/Tranliterator Signs/Cues Tests" and "Test Administrator Reads Test Aloud (in English)" accommodations during the administration of state tests that measure reading comprehension skills invalidates the results of the test.

6.3 Scale Score Frequency Distributions

The following tables present the frequency distributions of the scale scores from the first statewide operational administration of the North Carolina *EXTEND2* Science Tests.

Figure 3: Scale score frequency distribution for the 2008 Grade 5 *NCEXTEND2* Science Test

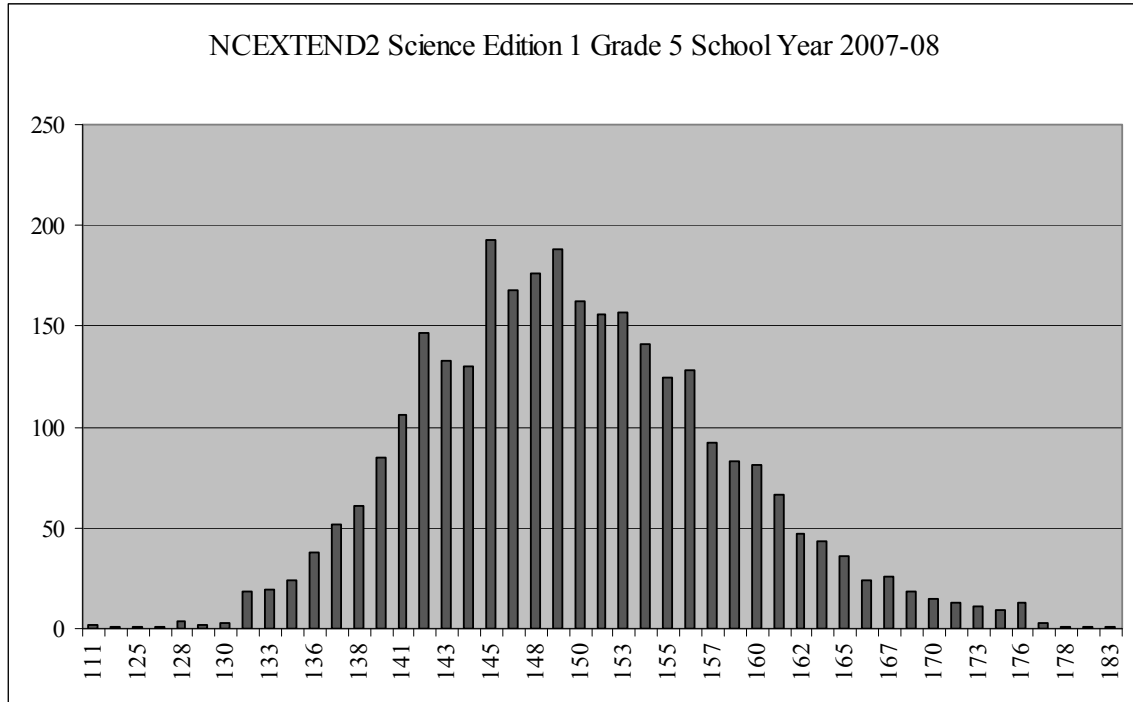
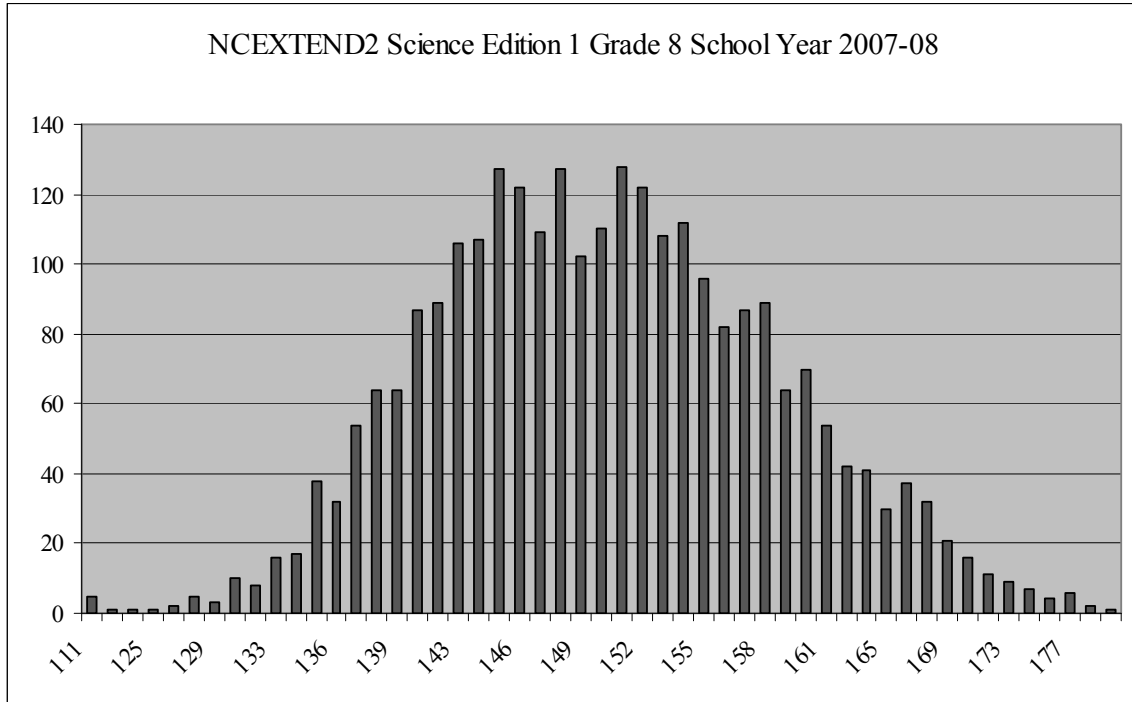


Figure 4: Scale score frequency distribution for the 2008 Grade 8 *NCEXTEND2* Science Test



6.4 Reliability of the North Carolina *EXTEND2* Science Tests

Reliability refers to the consistency of a measure when the testing procedure is repeated on a population of individuals or groups. Three broad categories of reliability coefficients are recognized as appropriate indices for establishing reliability in tests: (a) coefficients derived from the administration of parallel forms in independent testing sessions (alternate-form coefficients); (b) coefficients obtained by administration of the same instrument on separate occasions (test-retest or stability coefficients); and (c) coefficients based on the relationships among scores derived from individual items or subsets of the items within a test, all data accruing from a single administration of the test. The last coefficient, commonly known as the internal consistency coefficient (*Standards for Educational and Psychological Testing*, AERA, APA, NCME, 1985, p.27), is the coefficient used to establish reliability for the North Carolina Science Tests.

6.5 Internal Consistency of the North Carolina *EXTEND2* Science Tests

Internal-consistency reliability estimates examine the extent to which the test measures a single basic concept. One procedure for determining the internal consistency of a test is coefficient alpha (α). Coefficient alpha sets an upper limit to the reliability of tests constructed in terms of the domain sampling model. The formula for coefficient alpha is:

$$r_{xx} = \left[\frac{N}{N-1} \right] \left[\frac{S^2 - \sum s^2}{S^2} \right]$$

where

r_{xx} = Coefficient alpha

N = Number of items constituting the instrument

S^2 = Variance of the scores

$\sum s^2_i$ = The sum of the variances of the individual items that constitute this scale (Hatcher & Stepanski, 1994, *Using SAS System for Univariate and Multivariate Statistics*).

If any use is to be made of the information from a test, then test results must be reliable.

The following table presents the coefficient alpha indices averaged across forms by grade.

Table 13: Reliability indices for the 2008 operational year *NCEXTEND2* Tests of Science

Subject/ Grade	Coefficient Alpha
Science 5	0.722
Science 8	0.811

As noted above, the *NCEXTEND2* Tests of Science are reasonably reliable as a whole. In general the *NCEXTEND2* Tests of Science are less reliable than the general EOG Tests of Science. However, given the reduced number of items on the *NCEXTEND2* Tests of Science and the population of students assessed, this result is not surprising.

It is important to note that a similar degree of reliability extends across gender, ethnicity, primary language, and disability. The following tables provide a breakdown of coefficient alphas by grade and group for the tests given operationally during the 2007-2008 school year.

Table 14: Reliability indexes for 2008 operational year by gender

Subject/ Grade	Females	Males
Science 5	0.665	0.743
Science 8	0.788	0.818

Table 15: Reliability indexes for 2008 operational year by ethnicity

Grade	Asian	Black	Hispanic	Native Amer	Multi	White
Science 5	0.836	0.661	0.663	0.578	0.771	0.761
Science 8	0.753	0.757	0.769	0.756	0.819	0.842

6.6 Standard Error of Measurement

The information provided by the standard error of measurement for a given score is important because it assists in determining the accuracy of an examinee's obtained score. It allows a probabilistic statement to be made about an individual's test score. For example, if a score of 100 has an SEM of plus or minus two, then one can conclude that a student obtained a score of 100, which is accurate within plus or minus 2 points with a 68% confidence. In other words, a 68% confidence interval for a score of 100 is 98–102. If that student were to be retested, his or her score would be expected to be in the range of 98–102 about 68% of the time.

The classical standard errors of measurement (SEM) for raw scores on the *NCEXTEND2* Tests are provided below.

Table 16: Classical standard error of measurement by grade

Grade/Subject	Standard Error of Measurement
Science 5	4.48
Science 8	3.57

6.7 Decision Consistency and Accuracy of the *NCEXTEND2* Science Tests

Agreement coefficient is defined as the proportion of examinees consistently classified (in terms of proficient or non-proficient) on two parallel test administrations. As multiple test administrations is often impractical in applied situations, single administration methods of calculating agreement have been proposed (Huynh, 1976; Subkoviak, 1976). The tables below provide approximate values of the agreement coefficient and kappa coefficient, the proportion of consistent classification observed beyond that which would be expected based on chance. These values are approximated based off of the tables provided by Subkoviak, 1988.

Table 17: Approximate Agreement and Kappa Coefficient

Grade/Subject	Approximate Value of the Agreement Coefficient	Approximate Value of the Kappa Coefficient
Science 5	0.79	0.47
Science 8	0.82	0.58

Chapter Seven: Evidence of Validity

7.1 Evidence of Validity

The validity of a test is the degree to which evidence and theory support the interpretation of test scores. Validity provides a check on how well a test fulfills its function. For all forms of test development, the validity of the test is an issue to be addressed from the first stage of development through analysis and reporting of scores. The process of validation involves accumulating evidence to provide a sound scientific basis for the proposed test score interpretations. Those interpretations of test scores are evaluated rather than the test itself. Validation, when possible, should include several types of evidence and the quality of the evidence is of primary importance (AERA, APA, NCME, 1985). For the North Carolina *EXTEND2* EOG Tests of Science, evidence of validity is provided through content relevance, response processes, relationship of test scores to other external variables, and maintaining consistency in the testing environment.

7.2 Content Validity

Evidence of content validity begins with an explicit statement of the constructs or concepts being measured by the proposed test. The constructs or concepts measured by the North Carolina *EXTEND2* EOG Tests of Science are categorized by four basic strands: Nature of Science, Science as Inquiry, Science and Technology, and Science in Social and Personal Perspectives. These strands provide a context for teaching the content of the goals of the Standard Course of Study. Each item developed for the North Carolina *EXTEND2* EOG Tests of Science is written to measure one of those constructs.

All items developed for the *NCEXTEND2* EOG are done so to measure those concepts with particular focus on assessing students' ability to process information and engage in higher order thinking. The purpose of the test specification summary is to show the distribution of items across the curriculum. For test specification summaries, see Appendix B.

Content validity is further evidenced through the item development process. Almost all of the items are written by North Carolina teachers and other educators. Some of the science items were written under a contract with a North Carolina-based testing company to handle the logistics, but that contract specified that at least half of the items be written by teachers from North Carolina. Additionally, the items were all reviewed by North Carolina teachers. Many of the items on *NCEXTEND2* tests were unused items from the general assessment that were previously written and reviewed by North Carolina teachers and then revised to simplify language and reduce the number of foils by content and curriculum experts.

7.3 Instructional Validity

As a part of the test review process, DPI routinely administers questionnaires to teachers in order to evaluate the validity and appropriateness of the North Carolina **EXTEND2** EOG Tests of Science. Teachers are asked to respond to four statements. The statements for Grades 5 and 8 Science are provided as examples.

1. If the content of these forms DOES NOT reflect the goals and objectives of the curriculum as outline on the list of objectives, please explain below.
2. If the content of these forms DOES NOT reflect the goals and objectives of the curriculum as it is taught in your school or school system, please explain below.
3. If the content of these forms IS NOT balanced in relation to ethnicity, race, sex, socioeconomic status, or limited English proficiency, please explain below.
4. If you have any additional comments, please list below.

7.4 Criterion-Related Validity

Analysis of the relationship of test scores to variables external to the test provide another important source of validity evidence. External variables may include measures of some criteria that the test is expected to predict, as well as relationships to other tests hypothesized to measure the same constructs.

Criterion-related validity of a test indicates the effectiveness of a test in predicting an individual's behavior in a specific situation. The criterion for evaluating the performance of a test can be measured at the same time (concurrent validity) or at some later time (predictive validity). For the **NCEXTEND2**, teachers' judgments of student achievement, expected grade, and assigned achievement levels all serve as sources of evidence of concurrent validity. The Pearson correlation coefficient is used to provide a measure of association between the scale score and those variables listed above.

**Note: By comparison, the uncorrected correlation coefficient between SAT score and freshman year grades in college is variously reported as 0.35 to 0.55 (Camera & Echternacht, 2000).*

Table 18: Pearson Correlation Coefficients for the **NCEXTEND2** Science Tests (Grades 5 and 8)

Grade	5	8
Teacher Judgment of Achievement Level by Assigned Achievement Level	.275**	.230**

Teacher Judgment of Achievement by Expected Grade	.459**	.452**
Teacher Judgment of Achievement by Scale Score	.297**	.230**
Assigned Achievement Level by Expected Grade	.087**	-.019
Expected Grade by Scale Score	.096**	-.019

** Correlation is significant at .01 level (2-tailed).

The variables used in the tables above are as follows.

- **Teacher Judgment of Achievement:** Teachers were asked, for each student participating in the test, to evaluate the student’s absolute ability, external to the test, based on their knowledge of their students’ achievement. The categories that teachers could use correspond to the achievement level descriptors.
- **Scale Score:** The converted raw-score-to-scale-score value obtained by each examinee.
- **Assigned Achievement Level:** The achievement level assigned to a student based on his or her test score, based on the cut scores previously described on page 34.
- **Expected Grade:** Teachers were also asked to provide for each student the letter grade that they anticipated each student would receive at the end of the grade or course.

DPI found moderate correlations between *NCEXTEND2* scale scores and teachers’ judgment of student achievement. Equally important is to demonstrate that the test scores are *not* correlated with external variables that should not be related to student proficiency or should only be moderately correlated. The department also found generally low correlations among the test scores and variables external to the test such as gender, Migrant Status, Economic Disability Status (EDS), and Title I

The variables Migrant Status, Economic Disability Status (EDS), and Title I status were all coded so that not applicable was coded as 0. Thus, negative coefficients mean that student who were not, for example, migrant, did better than students who were migrant. Gender was coded as male or female; a negative coefficient for gender means that males did better than females.

Table 19: Tetrachoric Correlation Coefficients for the *NCEXTEND2* Science Tests (Grades 5 and 8)

Score by Grade	× Gender	× EDS	× Migrant	× Title I
Grade 5	-0.1237	-0.0782	-0.1947	0.0806
Grade 8	-0.1279	-0.1327	-0.0275	0.0511

Chapter Eight: Quality Control

Quality control procedures for the North Carolina testing program are implemented throughout all stages of testing. This includes quality control for test development, test administration, score analysis, and reporting.

8.1 Quality Control Prior to Test Administration

Once test forms have been assembled, they are reviewed by a panel of subject experts. Once the review panel has approved a test form, test forms are then configured to go through the printing process. Printers send a blue-lined form back to NCDPI Test Development staff to review and adjust if necessary. Once all test answer sheets and booklets are printed, the test project manager conducts a spot check of test booklets to ensure that all test pages are included and test items are in order.

8.2 Quality Control in Data Preparation and Test Administration

Student background information must be coded before testing begins. The school system may elect to either (1) pre-code the answer sheets, (2) direct the test administrator to code the Student Background Information, or (3) direct the students to code the Student Background Information. For the North Carolina multiple-choice tests, the school system may elect to precode some or all of the Student Background Information on SIDE 1 of the printed multiple-choice answer sheet. The precoded responses come from the schools' SIMS/NCWISE database. Precoded answer sheets provide schools with the opportunity to correct or update information in the SIMS/NCWISE database. In such cases, the test administrator ensures that the precoded information is accurate. The test administrator must know what information will be precoded on the student answer sheets to prepare for the test administration. Directions for instructing students to check the accuracy of these responses are located in test administrator manuals. All corrections for precoded responses are provided to a person designated by the school system test coordinator to make such corrections. The students and the test administrator must not change, alter, or erase precoding on students' answer sheets. To ensure that all students participate in the required tests and to eliminate duplications, students, regardless of whether they take the multiple-choice test or an alternate assessment, are required to complete the student background information on the answer sheets.

When tests and answer sheets are received by the local schools, they are kept in a locked, secure location. Class rosters are reviewed for accuracy by the test administrator to ensure that students receive their answer sheets. During test administration at the school level, proctors and test administrators circulate throughout the test facility (typically a classroom) to ensure that students are using the bubble sheets correctly. Once students have completed their tests, answer sheets are reviewed and, where appropriate, cleaned by local test coordinators (removal of stray marks, etc.).

8.3 Quality Control in Data Input

All answer sheets are then sent from individual schools to the Local Test Coordinator, where they are scanned in a secure facility. The use of a scanner provides the opportunity to program in a number of quality control mechanisms to ensure that errors overlooked in the manual check of data are identified and resolved. For example, if the answer sheet is unreadable by the scanner, the scanner stops the scan process until the error is resolved. In addition, if a student bubbles in two answers for the same question, the scan records the student's answer as a (*) indicating that the student has answered twice.

8.4 Quality Control of Test Scores

Once all tests are scanned, they are then sent through a secure system to the Regional Accountability Coordinators who checks to ensure that all schools in all LEAs have completed and returned student test scores. The Regional Accountability Coordinators also conduct a spot check of data and then send the data through a secure server to the North Carolina Department of Public Instruction Division of Accountability. Data are then imported into a file and cleaned. When a portion of the data are in, the NCDPI runs a CHECK KEYS program to flag areas where answer keys may need a second check. In addition, as data come into the NCDPI Division of Accountability Services, Reporting Section staff import and clean data to ensure that individual student files are complete.

8.5 Quality Control in Reporting

Scores can only be reported at the school level after the NCDPI issues a certification statement. This is to ensure that school-, district-, and state-level quality control procedures have been employed. The certification statement is issued by the NCDPI Division of Accountability. The following certification statement is an example:

“The department hereby certifies the accuracy of the data from the North Carolina end-of-course tests for Fall 2004 provided that all NCDPI-directed test administration guidelines, rules, procedures, and policies have been followed at the district and schools in conducting proper test administrations and in the generation of the data. The LEAs may generate the required reports for the end-of-course tests as this completes the certification process for the EOC tests for the Fall 2004 semester.”

Glossary of Key Terms

The terms below are defined by their application in this document and their common uses in the North Carolina Testing Program. Some of the terms refer to complex statistical procedures used in the process of test development. In an effort to avoid the use of excessive technical jargon, definitions have been simplified; however, they should not be considered exhaustive.

Accommodations	Changes made in the format or administration of the test to provide options to test takers who are unable to take the original test under standard test conditions.
Achievement levels	Descriptions of a test taker's competency in a particular area of knowledge or skill, usually defined as ordered categories on a continuum classified by broad ranges of performance.
Biserial correlation	The relationship between an item score (right or wrong) and a total test score.
Cut Scores	A specific point on a score scale, such that scores at or above that point are interpreted or acted upon differently from scores below that point.
Dimensionality	The extent to which a test item measures more than one ability.
Embedded test model	Using an operational test to field test new items or sections. The new items or sections are "embedded" into the new test and appear to examinees as being indistinguishable from the operational test.
Equivalent forms	Statistically insignificant differences between forms (i.e., the red form is not harder).
Field test	A collection of items to approximate how a test form will work. Statistics produced will be used in interpreting item behavior/performance and allow for the calibration of item parameters used in equating tests.
Foil counts	Number of examinees that endorse each foil (e.g., number who answer "A," number who answer "B," etc.)

Item response theory	A method of test item analysis that takes into account the ability of the examinee and determines characteristics of the item relative to other items in the test. The NCDPI uses the 3-parameter model, which provides slope, threshold, and asymptote.
Item Tryout	A collection of a limited number of items of a new type, a new format, or a new curriculum. Only a few forms are assembled to determine the performance of new items and not all objectives are tested.
Mantel-Haenszel	A statistical procedure that examines the differential item functioning (DIF) or the relationship between a score on an item and the different groups answering the item (e.g., gender, race). This procedure is used to examine individual items for bias.
Operational Test	Test is administered statewide with uniform procedures, full reporting of scores , and stakes for examinees and schools.
p-value	Difficulty of an item defined by using the proportion of examinees who answered an item correctly.
Parallel Forms	Covers the same curricular material as other forms
Percentile	The score on a test below which a given percentage of scores fall.
Pilot Test	Test is administered as if it were “the real thing” but has limited associated reporting or stakes for examinees or schools.
Quasi-equated	Item statistics are available for items that have been through item tryouts (although they could change after revisions) and field test forms are developed using this information to maintain similar difficulty levels to the extent possible.
Raw score	The unadjusted score on a test determined by counting the number of correct answers.
Scale score	A score to which raw scores are converted by numerical transformation. Scale scores allow for comparison of different forms of the test using the same scale.

Standard error of measurement	The standard deviation of an individual's observed scores usually estimated from group data.
Test Blueprint	The testing plan, which includes numbers of items from each objective to appear on test and arrangement of objectives.
Threshold	The point on the ability scale where the probability of a correct response is 50%. Threshold for an item of average difficulty is 0.00.
WINSCAN Program	Proprietary computer program that contains the test answer keys and files necessary to scan and score state multiple-choice tests. Student scores and local reports can be generated immediately using the program.

References

- Gregory, Robert J. (2000) *Psychological Testing: History, Principles, and Applications*. Needham Heights: Allyn & Bacon.
- Haladyna, T.M. & Downing, S.M. (1993). How Many Options is Enough for Multiple-Choice Test Items? *Educational and Psychological Measurement*, 53, 999-1010.
- Hinkle, Wiersma, Jurs, (1998). *Applied Statistics for the Behavioral Sciences*, Boston : Houghton Mifflin (pp.69-70)
- Marzano, R.J., Brandt, R.S., Hughes, C.S., Jones, B.F., Presseisen, B.Z., Stuart, C./, & Suhor, C. (1988). *Dimensions of Thinking*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Millman, J., and Greene, J. (1993). *The Specification and Development of Tests of Achievement and Ability*. In Robert Linn (ed.), *Educational Measurement* (pp. 335-366). Phoenix: American Council on Education and Oryx Press.
- Subkoviak, M. J. (1988). A practitioner's guide to computation and interpretation of reliability indices for mastery tests. *Journal of Educational Measurement*, 25, 47-55.

Additional Resources

- Anastasi, A. (1982). *Psychological Testing*. New York: Macmillan Publishing Company, Inc.
- Berk, R.A. (1984). *A Guide to Criterion-Referenced Test Construction*. Baltimore: The Johns Hopkins University Press.
- Berk, R.A. (1982). *Handbook of Methods for Detecting Test Bias*. Baltimore: The Johns Hopkins University Press.
- Camilli, G. & Shepard, L.A. (1994). *Methods for Identifying Biased Test Items*. Thousand Oaks, CA: Sage Publications, Inc.
- Dorans, N.J. & Holland, P.W. (1993). DIF Detection and description: Mantel-Haenszel and standardization. In P.W. Holland and H. Wainer (Eds.), *Differential Item Functioning* (pp 35-66). Hillsdale, NJ: Lawrence Erlbaum.
- Haladyna, T.M. (1994). *Developing and Validating Multiple-Choice Test Items*. Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.
- Hambleton, R.K. & Swaminathan, H. (1985). *Item Response Theory: Principles and Applications*. Kluwer-Nijhoff Publishing.

- Hambleton, R.K., Swaminathan, H., & Rogers, H.J. (1991). *Fundamentals of Item Response Theory*. Newbury Park, CA: Sage Publications, Inc.
- Holland, P.W. & Wainer, H. (1993). *Differential Item Functioning*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Joreskog, K.J. & Sorbom, D. (1986). *PRELIS: A program for multivariate data screening and data summarization*. Chicago, IL: Scientific Software, Inc.
- Joreskog, K.J. & Sorbom, D. (1988). *LISREL 7: A guide to the program and applications*. Chicago, IL: SPSS, Inc.
- Kubiszyn, T. & Borich, G. (1990). *Educational Testing and Measurement*. New York: HarperCollins Publishers.
- Muraki, E., Mislevy, R.J., & Bock, R.D. *PC-Bimain Manual*. (1991). Chicago, IL: Scientific Software, Inc.
- North Carolina Department of Public Instruction. (1993). *North Carolina End-of-Grade Testing Program: Background Information*. Raleigh, NC: Author.
- North Carolina Department of Public Instruction. (1996). *North Carolina Testing Code of Ethics*. Raleigh, NC: Author.
- Nunnally, J. (1978). *Psychometric Theory*. New York: McGraw-Hill Book Company.
- Rosenthal, R. & Rosnow, R.L. (1984). *Essentials of behavioral research: Methods and data analysis*. New York: McGraw-Hill Book Company.
- SAS Institute, Inc. (1985). *The FREQ Procedure*. In *SAS User's Guide: Statistics, Version 5 Edition*. Cary, NC: Author.
- Traub, R.E. (1994). *Reliability for the social sciences: Theory and applications*. Thousand Oaks, CA: Sage Publications, Inc.

Appendix A: Item Writing Guidelines

Content Guidelines

1. Items must be based on the goals and objectives outlined in the North Carolina *Standard Course of Study* in Science and written at the appropriate grade level.
2. To the extent possible, each item written should measure a single concept, principle, procedure, or competency.
3. Write items that measure important or significant material instead of trivial material.
4. Keep the testing vocabulary consistent with the expected grade level of students tested.
5. Avoid writing stems based on opinions.
6. Emphasize higher-level thinking skills using the taxonomy provided by the NCDPI.

Procedural Guidelines

7. Use the best answer format.
8. Avoid writing complex multiple-choice items.
9. Format the items vertically, not horizontally.
10. Avoid errors of grammar, abbreviations, punctuation, and spelling.
11. Minimize student reading time.
12. Avoid tricky or misleading items.
13. Avoid the use of contractions.
14. Avoid the use of first or second person.

Stem Construction Guidelines

15. Items are to be written in the question format.
16. Ensure that the directions written in the stems are clear and that the wording lets the students know exactly what is being tested.
17. Avoid excessive verbiage when writing the item stems.
18. Word the stems positively, avoiding any negative phrasing. The use of negatives such as NOT and EXCEPT is to be avoided.
19. Write the items so that the central idea and the phrasing are included in the stem instead of the foils.
20. Place the interrogative as close to the item foils as possible.

General Foil Development

21. Each item must contain four foils (A, B, C).
22. Order the answer choices in a logical order. Numbers should be listed in ascending or descending order.

23. Each item should contain foils that are independent and not overlapping.
24. All foils in an item should be homogeneous in content and length.
25. Do not use the following as foils: all of the above, none of the above, I don't know.
26. Word the foils positively, avoiding any negative phrasing. The use of negatives such as NOT and EXCEPT is to be avoided.
27. Avoid providing clues to the correct response. Avoid writing items with phrases in the stem (slang associations) that are repeated in the foils.
28. Also avoid including ridiculous options.
29. Avoid grammatical clues to the correct answer.
30. Avoid specific determiners since they are so extreme that they are seldom the correct response. To the extent possible, specific determiners such as ALWAYS, NEVER, TOTALLY, and ABSOLUTELY should not be used when writing items. Qualifiers such as *best*, *most likely*, *approximately*, etc. should be bold and italic.
31. The correct response for items written should be evenly balanced among the response options. For a 3-option multiple-choice item, each correct response should be located at each option position about 33 percent of the time.
32. Items should contain one and only one best (correct) answer.

Distractor Development

33. Use plausible distractors. The best (correct) answer must clearly be the best (correct) answer and the incorrect responses must clearly be inferior to the best (correct) answer. No distractor should be obviously wrong.
34. To the extent possible, use the common errors made by students as distractors. Give your reasoning for incorrect choices on the back of the item spec sheet.
35. Technically written phrases may be used, where appropriate, as plausible distractors.
36. True phrases that do not correctly respond to the stem may be used as plausible distractors where appropriate.
37. The use of humor should be avoided.

Appendix B. Test Specifications

Test Specification Outline

NCEXTEND2 EOG Grade 5 Science

Element	Comments
Purpose of the Test	The NCEXTEND2 End-of-Grade Test of Science at grade 5 is required to meet the federal <i>No Child Left Behind Act</i> of 2001. As a component of the North Carolina Annual Testing Program, this test falls subject to General Statute §115C-174.10 which states the purposes of North Carolina state mandated tests are “(i) to assure that all high school graduates possess those minimum skills and that knowledge thought necessary to function as a member of society, (ii) to provide a means of identifying strengths and weaknesses in the education process in order to improve instructional delivery, and (iii) to establish additional means for making the education system at the state, local, and school levels accountable to the public for results.”
Uses of the Test	Student scores are used in determining student progress and proficiency under state-mandated Student Accountability Standards at grade 5. According to State Board of Education policy, the standard for grade-level proficiency shall be a test score at Achievement Level III or above.
Curriculum Cycle	Test is based on the 2004 North Carolina Science <i>Standard Course of Study</i> .
Content of the Test	
Subject/Course & Grade	NCEXTEND2 Science Grade 5
Alignment	<p>The <i>NCEXTEND2</i> Alternate Assessment for EOG Science at grade 5 assesses the 2004 North Carolina <i>Standard Course of Study</i>. This test requires students to demonstrate knowledge of important principles and concepts, understand and interpret laboratory activities, and relate scientific information to everyday situations.</p> <p><i>NCEXTEND2</i> Alternate Assessment for EOG Science Grade 5 is an alternate assessment for students with disabilities who are working toward grade-level achievement but having continued difficulty in making progress in the same time frame as students without disabilities. The assessment is a multiple-choice test with fewer items that utilizes universal design principles to address</p>

	accessibility for students with disabilities. <i>NCEXTEND2</i> uses simplified language, and fewer test items and item responses (foils/answer choices) to assess students on grade-level content. <i>NCEXTEND2</i> provides access to the statewide testing program through a test design that utilizes a different format and permits the use of modified academic achievement standards (achievement levels). <i>NCEXTEND2</i> Alternate Assessment for EOG Science will be administered to students whose Individualized Education Program (IEP) designates <i>NCEXTEND2</i> as the appropriate assessment for end-of-grade science at grade 5.
Dimensionality	The construct of the test is unidimensional, requiring the reporting of a total score for the test.
Weighting	By Goal
Obj. not/indirectly measured	NA
Miscellaneous remarks	
Design	
Population	<p><i>NCEXTEND2</i> Alternate Assessment for EOG Science Grade 5 is an alternate assessment for students with disabilities who are working toward grade-level achievement but having continued difficulty in making progress in the same time frame as students without disabilities. <i>NCEXTEND2</i> Alternate Assessment for EOG Science will be administered to students whose Individualized Education Program (IEP) designates <i>NCEXTEND2</i> as the appropriate assessment for end-of-grade science at grade 5.</p> <p>To determine student participation in the <i>NCEXTEND2 EOG</i> Science Grade 5, the following eligibility requirements must be considered:</p> <ul style="list-style-type: none"> • The student must have a current IEP; • The student DOES NOT have a current 504 plan; • The student, if identified as limited English proficient (LEP), must also have a current IEP; • The student IS NOT identified as having a significant cognitive disability; • The student IS NOT receiving instruction in the SCOS through the Essences/Extensions; • The student’s progress in response to high-quality instruction is such that the student is not likely to achieve

	<p>grade level proficiency within the school year covered by the IEP;</p> <ul style="list-style-type: none"> • The student’s disability has precluded the student from achieving grade level proficiency, as demonstrated by objective evidence, (e.g., results from standardized state tests, IQ tests, achievement tests, aptitude tests, and psychological evaluations. It is the expectation that more than one objective measure would be used to assist in the evaluation of a student’s assessment placement.); and • The nature of the student’s disability may require assessments that are different in design.
Format	
Item type	Multiple choice: stem with three foils
Special item considerations	Items must be original and unique as well as free of bias (cultural, economic, gender, ethnic, or geographic). Distractors must be plausible and the language must be clear and precise.
Delivery Mode	
Mode	Paper & pencil, online pending
Accommodations	As written in a student’s IEP: Assistive Technology Devices, Braille, Braille Writer/Slate and Stylus, Cranmer Abacus, Dictation to Scribe, Home/Hospital Testing, Interpreter/Transliterator Signs/Cues Test, Keyboarding Devices, Large Print, Magnification Devices, Multiple Testing Sessions, One Test Item Per Page, Scheduled Extended Time, Student Reads Test Aloud to Self, Student Marks Answers in Test Book, Test Administrator Reads Test Aloud, Testing in Separate Room
Number of Items (Total)	60
Operational	60
Embedded	0
By section	15
Time Limits	None. Estimated time is 120 minutes.

Universal Design	Items and selections are reviewed for accessibility by all students, specifically students with disabilities and students with limited English proficiency.		
Item & Test Characteristics			
Item	Thinking Levels: Knowledge, Organizing, Applying, Analyzing, Generating, Integrating, Evaluating		
	NCSCS Goal/Objective		
	Difficulty Level(a priori): Easy, Medium, and Hard		
Test	Goal	Description of Category	Average Percentage
	1	The learner will conduct investigations to build an understanding of the interdependence of plants and animals.	25%-30%
	2	The learner will make observations and conduct investigations to build an understanding of landforms.	20%-25%
	3	The learner will conduct investigations and use appropriate technology to build an understanding of weather and climate.	20%-25%
	4	The learner will conduct investigations and use appropriate technologies to build an understanding of forces and motion in technological designs.	25%-30%
Cognitive Taxonomy	Dimensions of Thinking (Marzano et al.)		
Stimulus Materials	Not specified.		
Other Specifications			
Psychometric Properties			
P-value	.15<keep<.85 .85>reserve>.90 .10<reserve<.15		
Biserial Correlation	Keep >.25 Reserve >.15		
Slope	Keep >.7 .7 >Reserve >.5 *if applicable		

Asymptote	Keep<.35 .35<reserve<.45 *if applicable
Threshold	-2.5<keep<2.5 2.5≤reserve≤3.0 *if applicable
Dif Flags	.667<MH<1.5 not flagged
Minimum Reliability	.85 (multiple-choice)
Test Administration	
Guidelines & Procedures	Adhere to directions/script in Test Administrator's Manual
Materials	Four-function calculator with memory key.
Testing Window	<p>Last four weeks of the school year.</p> <ul style="list-style-type: none"> • The tests must be administered as early in the school day as the school schedule permits. • Afternoon administrations of the <i>NCEXTEND2</i> Alternate Assessment for EOG are prohibited except for students who receive certain testing accommodations (e.g., scheduled extended time or multiple testing sessions). • All students at the same grade level within a school must be administered the appropriate end-of grade test at the same time on the same day. • All grades 5 and 8 science test administrations must occur in one school day unless the student has approved test accommodations that require testing over more than one day. • Testing grade levels in two subjects (i.e., science and reading, science and mathematics, or reading and mathematics) on one day is prohibited. • Students at different grade levels cannot be administered tests in the same classroom. • The NCDPI requires that all students be allowed ample opportunity to complete the <i>NCEXTEND2</i> science tests at grades 5 and 8. At the school level, provisions must be made for students who will need time beyond that scheduled to complete the tests.
Scoring	

Methods	Scanned and scored locally (NCDPI provided software)
Raw Scores	0-48
Scale Scores	TBD
Standard Setting	
Achievement Level Ranges & Descriptors	<p style="text-align: center;">4 achievement levels</p> <p><u>Achievement Level I</u> Students performing at this level do not have sufficient mastery of knowledge and skills in this subject area to be successful at the next grade level.</p> <p><u>Achievement Level II</u> Students performing at this level demonstrate inconsistent mastery of knowledge and skills in this subject area and are minimally prepared to be successful at the next grade level.</p> <p><u>Achievement Level III</u> Students performing at this level consistently demonstrate mastery of grade level subject matter and skills and are well prepared for the next grade level.</p> <p><u>Achievement Level IV</u> Students performing at this level consistently perform in a superior manner clearly beyond that required to be proficient at grade level work.</p>
Method	Contrasting groups, item mapping
Reporting	
Levels of Reporting	Student, school, LEA, state
ABCs	Student report card, school report card, state report card
NCLB	None
Appropriate Use	Achievement Level III and above represent proficiency.
History of Development	

Committee Members	
Meeting Minutes	Not available

GOAL 1: The learner will conduct investigations to build an understanding of the interdependence of plants and animals.

25-30%

1.01	Describe and compare several common ecosystems (communities of organisms and their interaction with the environment).
1.02	Identify and analyze the functions of organisms within the population of the ecosystem: <ul style="list-style-type: none"> • Producers. • Consumers. • Decomposers.
1.03	Explain why an ecosystem can support a variety of organisms.
1.04	Discuss and determine the role of light, temperature, and soil composition in an ecosystem's capacity to support life.
1.05	Determine the interaction of organisms within an ecosystem.
1.06	Explain and evaluate some ways that humans affect ecosystems. <ul style="list-style-type: none"> • Habitat reduction due to development. • Pollulants. • Increased nutrients.
1.07	Determine how materials are recycled in nature.

GOAL 2: The learner will make observations and conduct investigations to build an understanding of landforms.

20-25%

2.01	Identify and analyze forces that cause change in landforms over time including. <ul style="list-style-type: none"> • Water and ice. • Wind. • Gravity.
2.02	Investigate and discuss the role of water cycle and how movement of water over and through the landscape helps shape landforms.
2.03	Discuss and consider the wearing away and movement of rock and soil in erosion and its importance in forming: <ul style="list-style-type: none"> • Canyons. • Valleys. • Meanders. • Tributaries.
2.04	Describe the deposition of eroded material and its importance in establishing landforms including: <ul style="list-style-type: none"> • Deltas. • Flood plains.
2.05	Discuss how the flow of water and the slope of the land affect erosion.

2.06	Identify and use models, maps, and aerial photographs as ways of representing landforms.
2.07	Discuss and analyze how humans influence erosion and deposition in local communities, including school grounds, as a result of: <ul style="list-style-type: none"> • Clearing land. • Planting vegetation. • Building dams.

GOAL 3: The learner will conduct investigations and use appropriate technology to build an understanding of weather and climate.

20-25%

3.01	Investigate the water cycle including the processes of: <ul style="list-style-type: none"> • Evaporation. • Condensation. • Precipitation. • Run-off.
3.02	Discuss and determine how the following are affected by predictable patterns of weather: <ul style="list-style-type: none"> • Temperature. • Wind direction and speed. • Precipitation. • Cloud cover. • Air pressure.
3.03	Describe and analyze the formation of various types of clouds and discuss their relation to weather systems.
3.04	Explain how global atmospheric movement patterns affect local weather.
3.05	Compile and use weather data to establish a climate record and reveal any trends.
3.06	Discuss and determine the influence of geography on weather and climate: <ul style="list-style-type: none"> • Mountains. • Sea breezes. • Water bodies.

GOAL 4: The learner will conduct investigations and use appropriate technologies to build an understanding of forces and motion in technological designs.

25-30%

4.01	Determine the motion of an object by following and measuring its position over time.
4.02	Evaluate how pushing or pulling forces can change the position and motion of an object.
4.03	Explain how energy is needed to make machines move. <ul style="list-style-type: none"> • Moving air. • Gravity.
4.04	Determine that an unbalanced force is needed to move an object or change its direction.

4.05	Determine factors that affect motion including: <ul style="list-style-type: none"> • Force. • Friction. • Inertia. • Momentum.
4.06	Build and use a model to solve a mechanical design problem. <ul style="list-style-type: none"> • Devise a test for the model. • Evaluate the results of test.
4.07	Determine how people use simple machines to solve problems.

Test Specification Outline
NCEXTEND2 EOG Grade 8 Science

Element	Comments
Purpose of the Test	The NCEXTEND2 End-of-Grade Test of Science at grade 8 is required to meet the federal <i>No Child Left Behind Act</i> of 2001. As a component of the North Carolina Annual Testing Program, this test falls subject to General Statute §115C-174.10 which states the purposes of North Carolina state mandated tests are “(i) to assure that all high school graduates possess those minimum skills and that knowledge thought necessary to function as a member of society, (ii) to provide a means of identifying strengths and weaknesses in the education process in order to improve instructional delivery, and (iii) to establish additional means for making the education system at the state, local, and school levels accountable to the public for results.”
Uses of the Test	Student scores are used in determining student progress and proficiency under state-mandated Student Accountability Standards at grade 8. According to State Board of Education policy, the standard for grade-level proficiency shall be a test score at Achievement Level III or above.
Curriculum Cycle	Test is based on the 2004 North Carolina Science <i>Standard Course of Study</i> .
Content of the Test	
Subject/Course & Grade	NCEXTEND2 Science Grade 8
Alignment	The <i>NCEXTEND2</i> Alternate Assessment for EOG Science at grade 8 assesses the 2004 North Carolina <i>Standard Course of Study</i> . This test requires students to demonstrate knowledge of important principles and concepts, understand and interpret laboratory activities, and relate scientific information to everyday situations.

	<p><i>NCEXTEND2</i> Alternate Assessment for EOG Science Grade 8 is an alternate assessment for students with disabilities who are working toward grade-level achievement but having continued difficulty in making progress in the same time frame as students without disabilities. The assessment is a multiple-choice test with fewer items that utilizes universal design principles to address accessibility for students with disabilities. <i>NCEXTEND2</i> uses simplified language, and fewer test items and item responses (foils/answer choices) to assess students on grade-level content. <i>NCEXTEND2</i> provides access to the statewide testing program through a test design that utilizes a different format and permits the use of modified academic achievement standards (achievement levels). <i>NCEXTEND2</i> Alternate Assessment for EOG Science will be administered to students whose Individualized Education Program (IEP) designates <i>NCEXTEND2</i> as the appropriate assessment for end-of-grade science at grade 8.</p>
Dimensionality	The construct of the test is unidimensional, requiring the reporting of a total score for the test.
Weighting	By Goal
Obj. not/indirectly measured	NA
Miscellaneous remarks	
Design	
Population	<p><i>NCEXTEND2</i> Alternate Assessment for EOG Science Grade 8 is an alternate assessment for students with disabilities who are working toward grade-level achievement but having continued difficulty in making progress in the same time frame as students without disabilities. <i>NCEXTEND2</i> Alternate Assessment for EOG Science will be administered to students whose Individualized Education Program (IEP) designates <i>NCEXTEND2</i> as the appropriate assessment for end-of-grade science at grade 8.</p> <p>To determine student participation in the <i>NCEXTEND2 EOG</i> Science Grade 8, the following eligibility requirements must be considered:</p> <ul style="list-style-type: none"> • The student must have a current IEP; • The student DOES NOT have a current 504 plan; • The student, if identified as limited English proficient

	<p>(LEP), must also have a current IEP;</p> <ul style="list-style-type: none"> • The student IS NOT identified as having a significant cognitive disability; • The student IS NOT receiving instruction in the SCOS through the Essences/Extensions; • The student’s progress in response to high-quality instruction is such that the student is not likely to achieve grade level proficiency within the school year covered by the IEP; • The student’s disability has precluded the student from achieving grade level proficiency, as demonstrated by objective evidence, (e.g., results from standardized state tests, IQ tests, achievement tests, aptitude tests, and psychological evaluations. It is the expectation that more than one objective measure would be used to assist in the evaluation of a student’s assessment placement.); and • The nature of the student’s disability may require assessments that are different in design.
Format	
Item type	Multiple choice: stem with three foils
Special item considerations	Items must be original and unique as well as free of bias (cultural, economic, gender, ethnic, or geographic). Distractors must be plausible and the language must be clear and precise.
Delivery Mode	
Mode	Paper & pencil, online pending
Accommodations	As written in a student’s IEP: Assistive Technology Devices, Braille, Braille Writer/Slate and Stylus, Cranmer Abacus, Dictation to Scribe, Home/Hospital Testing, Interpreter/Transliterator Signs/Cues Test, Keyboarding Devices, Large Print, Magnification Devices, Multiple Testing Sessions, One Test Item Per Page, Scheduled Extended Time, Student Reads Test Aloud to Self, Student Marks Answers in Test Book, Test Administrator Reads Test Aloud, Testing in Separate Room
Number of Items (Total)	60
Operational	60

Embedded	0																								
By section	12																								
Time Limits	None. Estimated time is 120 minutes.																								
Universal Design	Items and selections are reviewed for accessibility by all students, specifically students with disabilities and students with limited English proficiency.																								
Item & Test Characteristics																									
Item	Thinking Levels: Knowledge, Organizing, Applying, Analyzing, Generating, Integrating, Evaluating NCSCS Goal/Objective Difficulty Level(a priori): Easy, Medium, and Hard																								
Test	<table border="1"> <thead> <tr> <th>Goal</th> <th>Description of Category</th> <th>Average Percentage</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>The learner will design and conduct investigations to demonstrate an understanding of scientific inquiry.</td> <td>48%-52%</td> </tr> <tr> <td>2</td> <td>The learner will demonstrate an understanding of technological design.</td> <td>28%-32%</td> </tr> <tr> <td>3</td> <td>The learner will conduct investigations and utilize appropriate technologies and information systems to build an understanding of the hydrospace..</td> <td>28%-32%</td> </tr> <tr> <td>4</td> <td>The learner will conduct investigations and utilize technology and information systems to build an understanding of chemistry.</td> <td>25%-30%</td> </tr> <tr> <td>5</td> <td>The learner will conduct investigations and utilize appropriate technologies and information systems to build an understanding of evidence of evolution in organisms and landforms.</td> <td>20%-25%</td> </tr> <tr> <td>6</td> <td>The learner will conduct investigations, use models, simulations, and appropriate technologies and information systems to build an understanding of cell theory.</td> <td>8%-12%</td> </tr> <tr> <td>7</td> <td>The learner will conduct investigations, use models, simulations, and appropriate technologies and information systems to build an understanding of microbiology.</td> <td>8%-12%</td> </tr> </tbody> </table>	Goal	Description of Category	Average Percentage	1	The learner will design and conduct investigations to demonstrate an understanding of scientific inquiry.	48%-52%	2	The learner will demonstrate an understanding of technological design.	28%-32%	3	The learner will conduct investigations and utilize appropriate technologies and information systems to build an understanding of the hydrospace..	28%-32%	4	The learner will conduct investigations and utilize technology and information systems to build an understanding of chemistry.	25%-30%	5	The learner will conduct investigations and utilize appropriate technologies and information systems to build an understanding of evidence of evolution in organisms and landforms.	20%-25%	6	The learner will conduct investigations, use models, simulations, and appropriate technologies and information systems to build an understanding of cell theory.	8%-12%	7	The learner will conduct investigations, use models, simulations, and appropriate technologies and information systems to build an understanding of microbiology.	8%-12%
Goal	Description of Category	Average Percentage																							
1	The learner will design and conduct investigations to demonstrate an understanding of scientific inquiry.	48%-52%																							
2	The learner will demonstrate an understanding of technological design.	28%-32%																							
3	The learner will conduct investigations and utilize appropriate technologies and information systems to build an understanding of the hydrospace..	28%-32%																							
4	The learner will conduct investigations and utilize technology and information systems to build an understanding of chemistry.	25%-30%																							
5	The learner will conduct investigations and utilize appropriate technologies and information systems to build an understanding of evidence of evolution in organisms and landforms.	20%-25%																							
6	The learner will conduct investigations, use models, simulations, and appropriate technologies and information systems to build an understanding of cell theory.	8%-12%																							
7	The learner will conduct investigations, use models, simulations, and appropriate technologies and information systems to build an understanding of microbiology.	8%-12%																							

Cognitive Taxonomy	Dimensions of Thinking (Marzano et al.)
Stimulus Materials	Not specified.
Other Specifications	
Psychometric Properties	
P-value	.15<keep<.85 .85>reserve>.90 .10<reserve<.15
Biserial Correlation	Keep >.25 Reserve >.15
Slope	Keep >.7 .7 >Reserve >.5 *if applicable
Asymptote	Keep<.35 .35<reserve<.45 *if applicable
Threshold	-2.5<keep<2.5 2.5≤reserve≤3.0 *if applicable
Dif Flags	.667<MH<1.5 not flagged
Minimum Reliability	.85 (multiple-choice)
Test Administration	
Guidelines & Procedures	Adhere to directions/script in Test Administrator’s Manual
Materials	Any four-function calculator with a square root function, y^x , $\Pi(\pi)$, and algebraic logic. A periodic table.
Testing Window	Last four weeks of the school year. <ul style="list-style-type: none"> • The tests must be administered as early in the school day as the school schedule permits. • Afternoon administrations of the <i>NCEXTEND2</i> Alternate Assessment for EOG are prohibited except for students who receive certain testing accommodations (e.g.,

	<p>scheduled extended time or multiple testing sessions).</p> <ul style="list-style-type: none"> • All students at the same grade level within a school must be administered the appropriate end-of grade test at the same time on the same day. • All grades 5 and 8 science test administrations must occur in one school day unless the student has approved test accommodations that require testing over more than one day. • Testing grade levels in two subjects (i.e., science and reading, science and mathematics, or reading and mathematics) on one day is prohibited. • Students at different grade levels cannot be administered tests in the same classroom. • The NCDPI requires that all students be allowed ample opportunity to complete the <i>NCEXTEND2</i> science tests at grades 5 and 8. At the school level, provisions must be made for students who will need time beyond that scheduled to complete the tests.
Scoring	
Methods	Scanned and scored locally (NCDPI provided software)
Raw Scores	0-48
Scale Scores	TBD
Standard Setting	
Achievement Level Ranges & Descriptors	<p style="text-align: center;">4 achievement levels</p> <p><u>Achievement Level I</u> Students performing at this level do not have sufficient mastery of knowledge and skills in this subject area to be successful at the next grade level.</p> <p><u>Achievement Level II</u> Students performing at this level demonstrate inconsistent mastery of knowledge and skills in this subject area and are minimally prepared to be successful at the next grade level.</p> <p><u>Achievement Level III</u> Students performing at this level consistently demonstrate mastery of grade level subject matter and skills and are well prepared for the next grade level.</p> <p><u>Achievement Level IV</u> Students performing at this level consistently perform in a</p>

	superior manner clearly beyond that required to be proficient at grade level work.
Method	Contrasting groups, item mapping
Reporting	
Levels of Reporting	Student, school, LEA, state
ABCs	Student report card, school report card, state report card
NCLB	None
Appropriate Use	Achievement Level III and above represent proficiency.
History of Development	
Committee Members	
Meeting Minutes	Not available

GOAL 1: The learner will design and conduct investigations to demonstrate an understanding of scientific inquiry. 50%	
1.01	Identify and create questions and hypotheses that can be answered through scientific investigations.
1.02	Develop appropriate experimental procedures for: <ul style="list-style-type: none"> • Given questions. • Student generated questions.
1.03	Apply safety procedures in the laboratory and in field studies: <ul style="list-style-type: none"> • Recognize potential hazards. • Safely manipulate materials and equipment. • Conduct appropriate procedures.

1.04	Analyze variables in scientific investigations: <ul style="list-style-type: none"> • Identify dependent and independent. • Use of a control. • Manipulate. • Describe relationships between. • Define operationally.
1.05	Analyze evidence to: <ul style="list-style-type: none"> • Explain observations. • Make inferences and predictions. • Develop the relationship between evidence and explanation.
1.06	Use mathematics to gather, organize, and present quantitative data resulting from scientific investigations: <ul style="list-style-type: none"> • Measurement. • Analysis of data. • Graphing. • Prediction models.
1.07	Prepare models and/or computer simulations to: <ul style="list-style-type: none"> • Test hypotheses. • Evaluate how data fit. • Make predictions.
1.08	Use oral and written language to: <ul style="list-style-type: none"> • Communicate findings. • Defend conclusions of scientific investigations. • Describe strengths and weaknesses of claims, arguments, and/or data.
1.09	Use technologies and information systems to: <ul style="list-style-type: none"> • Research. • Gather and analyze data. • Visualize data. • Disseminate findings to others.
1.10	Analyze and evaluate information from a scientifically literate viewpoint by reading, hearing, and/or viewing: <ul style="list-style-type: none"> • Scientific text. • Articles. • Events in the popular press.

GOAL 2: The learner will demonstrate an understanding of technological design.

30%

2.01	Explore evidence that “technology” has many definitions. <ul style="list-style-type: none"> • Artifact or hardware. • Methodology or technique. • System or production. • Social-technical system.
------	--

2.02	Use information systems to: <ul style="list-style-type: none"> Identify scientific needs, human needs, or problems that are subject to technological solution. Locate resources to obtain and test ideas.
2.03	Evaluate technological designs for: <ul style="list-style-type: none"> Application of scientific principles. Risks and benefits. Constraints of design. Consistent testing protocols.
2.04	Apply tenets of technological design to make informed consumer decisions about: <ul style="list-style-type: none"> Products. Processes. Systems.

GOAL 3: The learner will conduct investigations and utilize appropriate technologies and information systems to build an understanding of the hydrosphere. 28-32%	
3.01	Analyze the unique properties of water including: <ul style="list-style-type: none"> Universal solvent. Cohesion and adhesion. Polarity. Density and buoyancy. Specific Heat.
3.02	Explain the structure of the hydrosphere including: <ul style="list-style-type: none"> Water distribution on earth. Local river basin. Local water availability.
3.03	Evaluate evidence that Earth's oceans are a reservoir of nutrients, minerals, dissolved gases, and life forms: <ul style="list-style-type: none"> Estuaries. Marine ecosystems. Upwelling. Behavior of gases in the marine environment. Value and sustainability of marine resources. Deep ocean technology and understandings gained.
3.04	Describe how terrestrial and aquatic food webs are interconnected.
3.05	Analyze hydrospheric data over time to predict the health of a water system including: <ul style="list-style-type: none"> Temperature. Dissolved oxygen. pH. Nitrates. Turbidity. Bio-indicators.

3.06	Evaluate technologies and information systems used to monitor the hydrosphere.
3.07	Describe how humans affect the quality of water: <ul style="list-style-type: none"> • Point and non-point sources of water pollution in North Carolina. • Possible effects of excess nutrients in North Carolina waters. • Economic trade-offs. • Local water issues.
3.08	Recognize that the good health of environments and organisms requires: <ul style="list-style-type: none"> • Monitoring of the hydrosphere. • Water quality standards. • Methods of water treatment. • Maintaining safe water quality. • Stewardship.

GOAL 4: The learner will conduct investigations and utilize technology and information systems to build an understanding of chemistry.

25-30%

4.01	Understand that both naturally occurring and synthetic substances are chemicals.
4.02	Evaluate evidence that elements combine in a multitude of ways to produce compounds that account for all living and nonliving substances.
4.03	Explain how the periodic table is a model for: <ul style="list-style-type: none"> • Classifying elements. • Identifying the properties of elements.
4.04	Describe the suitability of materials for use in technological design: <ul style="list-style-type: none"> • Electrical conductivity. • Density. • Magnetism. • Solubility. • Malleability.
4.05	Identify substances based on a characteristic physical properties: <ul style="list-style-type: none"> • Density. • Boiling/melting points. • Solubility. • Chemical reactivity. • Specific heat.
4.06	Describe and measure quantities related to chemical/physical changes within a system: <ul style="list-style-type: none"> • Temperature. • Volume. • Mass. • Precipitate. • Gas production.

4.07	<p>Identify evidence supporting the law of conservation of matter.</p> <ul style="list-style-type: none"> • During an ordinary chemical reaction matter cannot be created or destroyed. • In a chemical reaction, the total mass of the reactants equals the total mass of the products.
4.08	<p>Identify evidence that some chemicals may contribute to human health conditions including:</p> <ul style="list-style-type: none"> • Cancer. • Autoimmune disease. • Birth defects. • Heart disease. • Diabetes. • Learning and behavioral disorders. • Kidney disease. • Asthma.
4.09	<p>Describe factors that determine the effects a chemical has on a living organism including:</p> <ul style="list-style-type: none"> • Exposure. • Potency. • Dose and the resultant concentration of chemical in the organism. • Individual susceptibility. • Possible means to eliminate or reduce effects.
4.10	<p>Describe risks and benefits of chemicals including:</p> <ul style="list-style-type: none"> • Medicines. • Food preservations. • Crop yield. • Sanitation.

GOAL 5: The learner will conduct investigations and utilize appropriate technologies and information systems to build an understanding of evidence of evolution in organisms and landforms.

20-25%

5.01	<p>Interpret ways in which rocks, fossils, and ice cores record Earth’s geologic history and the evolution of life including:</p> <ul style="list-style-type: none"> • Geologic time scale. • Index fossils. • Law of Superposition. • Unconformity. • Evidence for climate change. • Extinction of species. • Catastrophic events.
5.02	<p>Correlate evolutionary theories and processes:</p> <ul style="list-style-type: none"> • Biological. • Geological. • Technological.

5.03	Examine evidence that the geologic evolution has had significant global impact including: <ul style="list-style-type: none"> • Distribution of living things. • Major geological events. • Mechanical and chemical weathering.
5.04	Analyze satellite imagery as a method to monitor Earth from space: <ul style="list-style-type: none"> • Spectral analysis. • Reflectance curves.
5.05	Use maps, ground truthing and remote sensing to make predictions regarding: <ul style="list-style-type: none"> • Changes over time. • Land use. • Urban sprawl. • Resource management.

GOAL 6: The learner will conduct investigations, use models, simulations, and appropriate technologies and information systems to build an understanding of cell theory.

8-12%

6.01	Describe cell theory: <ul style="list-style-type: none"> • All living things are composed of cells. • Cells provide structure and carry on major functions to sustain life. • Some organisms are single cell; other organisms, including humans, are multi-cellular. • Cell function is similar in all living things.
6.02	Analyze structures, functions, and processes within animal cells for: <ul style="list-style-type: none"> • Capture and release of energy. • Feedback information. • Dispose of wastes. • Reproduction. • Movement. • Specialized needs.
6.03	Compare life functions of protists: <ul style="list-style-type: none"> • <i>Euglena</i>. • <i>Amoeba</i>. • <i>Paramecium</i>. • <i>Volvox</i>.
6.04	Conclude that animal cells carry on complex chemical processes to balance the needs of the organism. <ul style="list-style-type: none"> • Cells grow and divide to produce more cells. • Cells take in nutrients to make the energy for the work cells do. • Cells take in materials that a cell or an organism needs.

GOAL 7: The learner will conduct investigations, use models, simulations, and appropriate technologies and information systems to build an understanding of microbiology.

8-12%

7.01	<p>Compare and contrast microbes:</p> <ul style="list-style-type: none"> • Size, shape, structure • Whether they are living cells.
7.02	<p>Describe diseases caused by microscopic biological hazards including:</p> <ul style="list-style-type: none"> • Viruses. • Bacteria. • Parasites. • Contagions. • Mutagens.
7.03	<p>Analyze data to determine trends or patterns to determine how an infectious disease may spread including:</p> <ul style="list-style-type: none"> • Carriers. • Vectors. • Conditions conducive to disease. • Calculate reproductive potential of bacteria.
7.04	<p>Evaluate the human attempt to reduce the risk of and treatments for microbial infections including:</p> <ul style="list-style-type: none"> • Solutions with antimicrobial properties. • Antibiotic treatment. • Research.
7.05	<p>Investigate aspects of biotechnology including:</p> <ul style="list-style-type: none"> • Specific genetic information available. • Careers. • Economic benefits to North Carolina. • Ethical issues. • Impact for agriculture.

Appendix C: Scale Score Frequency Distribution Tables

Grade 5 *NCEXTEND2* Science Scale Score Frequency Distribution (2007-08)

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
111	2	0.07	2	0.07
123	1	0.03	3	0.10
125	1	0.03	4	0.13
126	1	0.03	5	0.17
128	4	0.13	9	0.30
129	2	0.07	11	0.37
130	3	0.10	14	0.47
132	18	0.60	32	1.07
133	19	0.63	51	1.70
134	24	0.80	75	2.50
136	38	1.26	113	3.76
137	52	1.73	165	5.49
138	61	2.03	226	7.52
139	85	2.83	311	10.35
141	106	3.53	417	13.88
142	147	4.89	564	18.77
143	133	4.43	697	23.20
144	130	4.33	827	27.53
145	193	6.42	1020	33.95
147	168	5.59	1188	39.55
148	176	5.86	1364	45.41
149	188	6.26	1552	51.66
150	162	5.39	1714	57.06
151	156	5.19	1870	62.25
153	157	5.23	2027	67.48
154	141	4.69	2168	72.17
155	125	4.16	2293	76.33
156	128	4.26	2421	80.59
157	92	3.06	2513	83.66
159	83	2.76	2596	86.42
160	81	2.70	2677	89.11
161	66	2.20	2743	91.31
162	47	1.56	2790	92.88
164	43	1.43	2833	94.31
165	36	1.20	2869	95.51
166	24	0.80	2893	96.30
167	26	0.87	2919	97.17
169	18	0.60	2937	97.77
170	15	0.50	2952	98.27
171	13	0.43	2965	98.70

173	11	0.37	2976	99.07
174	9	0.30	2985	99.37
176	13	0.43	2998	99.80
177	3	0.10	3001	99.90
178	1	0.03	3002	99.93
180	1	0.03	3003	99.97
183	1	0.03	3004	100.00

Grade 8 *NCEXTEND2* Science Scale Score Frequency Distribution (2007-08)

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
111	5	0.19	5	0.19
123	1	0.04	6	0.22
124	1	0.04	7	0.26
125	1	0.04	8	0.30
127	2	0.07	10	0.37
128	5	0.19	15	0.56
129	3	0.11	18	0.67
130	10	0.37	28	1.05
132	8	0.30	36	1.34
133	16	0.60	52	1.94
134	17	0.63	69	2.58
135	38	1.42	107	4.00
136	32	1.19	139	5.19
137	54	2.02	193	7.21
138	64	2.39	257	9.60
139	64	2.39	321	11.99
140	87	3.25	408	15.24
142	89	3.32	497	18.56
143	106	3.96	603	22.52
144	107	4.00	710	26.51
145	127	4.74	837	31.25
146	122	4.56	959	35.81
147	109	4.07	1068	39.88
148	127	4.74	1195	44.62
149	102	3.81	1297	48.43
150	110	4.11	1407	52.54
151	128	4.78	1535	57.32
152	122	4.56	1657	61.87
153	108	4.03	1765	65.91
154	112	4.18	1877	70.09
155	96	3.58	1973	73.67
156	82	3.06	2055	76.74
157	87	3.25	2142	79.99
158	89	3.32	2231	83.31
159	64	2.39	2295	85.70
160	70	2.61	2365	88.31
161	54	2.02	2419	90.33
163	42	1.57	2461	91.90
164	41	1.53	2502	93.43
165	30	1.12	2532	94.55
166	37	1.38	2569	95.93
167	32	1.19	2601	97.12

169	21	0.78	2622	97.91
170	16	0.60	2638	98.51
171	11	0.41	2649	98.92
173	9	0.34	2658	99.25
174	7	0.26	2665	99.51
176	4	0.15	2669	99.66
177	6	0.22	2675	99.89
179	2	0.07	2677	99.96
186	1	0.04	2678	100.00

Appendix D: Testing Code of Ethics

Testing Code of Ethics (16 NCAC 6D .0306)

Testing Code of Ethics

Introduction

In North Carolina, standardized testing is an integral part of the educational experience of all students. When properly administered and interpreted, test results provide an independent, uniform source of reliable and valid information, which enables:

- *students* to know the extent to which they have mastered expected knowledge and skills and how they compare to others;
- *parents* to know if their children are acquiring the knowledge and skills needed to succeed in a highly competitive job market;
- *teachers* to know if their students have mastered grade-level knowledge and skills in the curriculum and, if not, what weaknesses need to be addressed;
- *community leaders and lawmakers* to know if students in North Carolina schools are improving their performance over time and how the students compare with students from other states or the nation; and
- *citizens* to assess the performance of the public schools.

Testing should be conducted in a fair and ethical manner, which includes:

Security

- assuring adequate security of the testing materials before, during, and after testing and during scoring
- assuring student confidentiality

Preparation

- teaching the tested curriculum and test-preparation skills
- training staff in appropriate testing practices and procedures
- providing an appropriate atmosphere

Administration

- developing a local policy for the implementation of fair and ethical testing practices and for resolving questions concerning those practices
- assuring that all students who should be tested are tested
- utilizing tests which are developmentally appropriate
- utilizing tests only for the purposes for which they were designed

Scoring, Analysis and Reporting

- interpreting test results to the appropriate audience
- providing adequate data analyses to guide curriculum implementation and improvement

Because standardized tests provide only one valuable piece of information, such information should be used in conjunction with all other available information known about a student to assist in improving student learning. The administration of tests required by applicable statutes and the use of student data for personnel/program decisions shall comply with the *Testing Code of Ethics* (16 NCAC 6D .0306), which is

printed on the next three pages.

Testing Code of Ethics (16 NCAC 6D .0306)

.0306 TESTING CODE OF ETHICS

(a) This Rule shall apply to all public school employees who are involved in the state testing program.

(b) The superintendent or superintendent's designee shall develop local policies and procedures to ensure maximum test security in coordination with the policies and procedures developed by the test publisher. The principal shall ensure test security within the school building.

(1) The principal shall store test materials in a secure, locked area. The principal shall allow test materials to be distributed immediately prior to the test administration. Before each test administration, the building level test coordinator shall accurately count and distribute test materials. Immediately after each test administration, the building level test coordinator shall collect, count, and return all test materials to the secure, locked storage area.

(2) "Access" to test materials by school personnel means handling the materials but does not include reviewing tests or analyzing test items. The superintendent or superintendent's designee shall designate the personnel who are authorized to have access to test materials.

(3) Persons who have access to secure test materials shall not use those materials for personal gain.

(4) No person may copy, reproduce, or paraphrase in any manner or for any reason the test materials without the express written consent of the test publisher.

(5) The superintendent or superintendent's designee shall instruct personnel who are responsible for the testing program in testing administration procedures. This instruction shall include test administrations that require procedural modifications and shall emphasize the need to follow the directions outlined by the test publisher.

(6) Any person who learns of any breach of security, loss of materials, failure to account for materials, or any other deviation from required security procedures shall immediately report that information to the principal, building level test coordinator, school system test coordinator, and state level test coordinator.

(c) Preparation for testing.

(1) The superintendent shall ensure that school system test coordinators:

(A) secure necessary materials;

(B) plan and implement training for building level test coordinators, test administrators, and proctors;

(C) ensure that each building level test coordinator and test administrator is trained in the implementation

of procedural modifications used during test administrations; and

(D) in conjunction with program administrators, ensure that the need for test modifications is documented and that modifications are limited to the specific need.

(2) The principal shall ensure that the building level test coordinators:

(A) maintain test security and accountability of test materials;

(B) identify and train personnel, proctors, and backup personnel for test administrations; and

(C) encourage a positive atmosphere for testing.

(3) Test administrators shall be school personnel who have professional training in education and the state testing program.

(4) Teachers shall provide instruction that meets or exceeds the standard course of study to meet the needs of the specific students in the class. Teachers may help students improve test-taking skills by:

- (A) helping students become familiar with test formats using curricular content;
- (B) teaching students test-taking strategies and providing practice sessions;
- (C) helping students learn ways of preparing to take tests; and
- (D) using resource materials such as test questions from test item banks, testlets and linking documents in instruction and test preparation.

(d) Test administration.

(1) The superintendent or superintendent's designee shall:

- (A) assure that each school establishes procedures to ensure that all test administrators comply with test publisher guidelines;
- (B) inform the local board of education of any breach of this code of ethics; and
- (C) inform building level administrators of their responsibilities.

(2) The principal shall:

- (A) assure that school personnel know the content of state and local testing policies;
- (B) implement the school system's testing policies and procedures and establish any needed school policies and procedures to assure that all eligible students are tested fairly;
- (C) assign trained proctors to test administrations; and
- (D) report all testing irregularities to the school system test coordinator.

(3) Test administrators shall:

- (A) administer tests according to the directions in the administration manual and any subsequent updates developed by the test publisher;
- (B) administer tests to all eligible students;
- (C) report all testing irregularities to the school system test coordinator; and
- (D) provide a positive test-taking climate.

(4) Proctors shall serve as additional monitors to help the test administrator assure that testing occurs fairly.

(e) Scoring. The school system test coordinator shall:

(1) ensure that each test is scored according to the procedures and guidelines defined for the test by the test publisher;

(2) maintain quality control during the entire scoring process, which consists of handling and editing documents, scanning answer documents, and producing electronic files and reports. Quality control shall address at a minimum accuracy and scoring consistency.

(3) maintain security of tests and data files at all times, including:

- (A) protecting the confidentiality of students at all times when publicizing test results; and
- (B) maintaining test security of answer keys and item-specific scoring rubrics.

(f) Analysis and reporting. Educators shall use test scores appropriately. This means that the educator recognizes that a test score is only one piece of information and must be interpreted together with other scores and indicators. Test data help educators understand educational patterns and practices. The superintendent shall ensure that school personnel analyze and report test data ethically and within the limitations described in this paragraph.

(1) Educators shall release test scores to students, parents, legal guardians, teachers, and the media with interpretive materials as needed.

(2) Staff development relating to testing must enable personnel to respond knowledgeably to questions related to testing, including the tests, scores, scoring procedures, and other interpretive materials.

(3) Items and associated materials on a secure test shall not be in the public domain. Only items that are within the public domain may be used for item analysis.

(4) Educators shall maintain the confidentiality of individual students. Publicizing test scores that contain the names of individual students is unethical.

(5) Data analysis of test scores for decision-making purposes shall be based upon:

(A) disaggregation of data based upon student demographics and other collected variables;

(B) examination of grading practices in relation to test scores; and

(C) examination of growth trends and goal summary reports for state-mandated tests.

(g) Unethical testing practices include, but are not limited to, the following practices:

(1) encouraging students to be absent the day of testing;

(2) encouraging students not to do their best because of the purposes of the test;

(3) using secure test items or modified secure test items for instruction;

(4) changing student responses at any time;

(5) interpreting, explaining, or paraphrasing the test directions or the test items;

(6) reclassifying students solely for the purpose of avoiding state testing;

(7) not testing all eligible students;

(8) failing to provide needed modifications during testing, if available;

(9) modifying scoring programs including answer keys, equating files, and lookup tables;

(10) modifying student records solely for the purpose of raising test scores;

(11) using a single test score to make individual decisions; and

(12) misleading the public concerning the results and interpretations of test data.

(h) In the event of a violation of this Rule, the SBE may, in accordance with the contested case provisions of Chapter 150B of the General Statutes, impose any one or more of the following sanctions:

(1) withhold ABCs incentive awards from individuals or from all eligible staff in a school;

(2) file a civil action against the person or persons responsible for the violation for copyright infringement or for any other available cause of action;

(3) seek criminal prosecution of the person or persons responsible for the violation; and

(4) in accordance with the provisions of 16 NCAC 6C .0312, suspend or revoke the professional license of the person or persons responsible for the violation.

History Note: Authority G.S. 115C-12(9)c.; 115C-81(b)(4);

Eff. November 1, 1997;

Amended Eff. August 1, 2000.

Appendix E: Internal Standard Setting Meeting Attendees

***NCEXTEND2* Science/Reading Standard Setting
Sept 22, 2008
NCDPI Internal meeting**

<u>Participant</u>	<u>Division</u>	<u>Gender</u>	<u>Ethnicity</u>
Tammy Howard	Acct.	F	W
Melinda Taylor	Acct.	F	W
Nadine McBride	Acct.	F	W
Nancy Carolan	Acct.	F	W
Tom Winton	EC	M	W
Sheila Garner Brown	TOPS	F	W
Pam Biggs	Acct.	F	W
Beverly Vance	CI&T	F	B
Sarah McManus	Acct.	F	B
Joanne Marino	CI&T	F	W
Lou Fabrizio	Acct.	M	W
Gary L. Williamson	Acct.	M	W
Patricia Chalmers	CI&T	F	B

Appendix F: Achievement Level Descriptors

NCEXTEND2 EOG Grade 5 Science—Achievement Level Descriptors

Achievement Level I

Students performing at this level demonstrate insufficient knowledge and skills in this subject area to be successful at the next grade level.

Students demonstrate insufficient understanding of the diversity within ecosystems, the concept that basic landforms change, weather and climate, and the concepts of force and motion.

Achievement Level II

Students performing at this level demonstrate inconsistent knowledge and skills in this subject area and are minimally prepared to be successful at the next grade level.

Students demonstrate inconsistent understanding of the relationships within ecosystems, the components of the processes that change basic landforms, some factors influencing weather and climate, and the relationship between force and motion.

Achievement Level III

Students performing at this level consistently demonstrate sufficient knowledge of grade level subject matter and skills and are prepared for the next grade level.

Students demonstrate a sufficient understanding of the interdependence within ecosystems, the processes that change basic landforms, factors influencing weather and climate, and forces and motion in basic technological designs.

Achievement Level IV

Students performing at this level consistently demonstrate content mastery clearly beyond that required to be proficient and are well prepared for the next grade level.

Students demonstrate a considerable understanding of the interdependence within ecosystems as well as the ability to analyze the relationships within, the processes that change basic landforms while considering possible outcomes, the factors influencing weather and climate as well as interpreting how changes would affect the system, and forces and motion in basic technological designs as well as applying the concepts to possible situations.

NCEXTEND2 EOG Grade 8 Science—Achievement Level Descriptors

Achievement Level I

Students performing at this level demonstrate insufficient knowledge and skills in this subject area to be successful at the next grade level.

Students demonstrate insufficient understanding of some basic principles of scientific inquiry and technological design; basic characteristics of water; chemical changes in matter; some factors affecting biological and geological evolution; some basic structures of an animal cell; micro-organisms that cause diseases.

Achievement Level II

Students performing at this level demonstrate inconsistent knowledge and skills in this subject area and are minimally prepared to be successful at the next grade level.

Students demonstrate inconsistent understanding of basic principles and methodologies of scientific inquiry and technological design; the distribution, use, and properties of water and the water cycle; some properties of pure substances and the recognition of chemical changes in matter; factors affecting biological and geological evolution which cause change over time; the basic structures and some functions of an animal cell and some single-celled organisms; simple relationships between micro-organisms and diseases, and basic applications of biotechnology.

Achievement Level III

Students performing at this level consistently demonstrate sufficient knowledge of grade level subject matter and skills and are prepared for the next grade level.

Students demonstrate a sufficient understanding of principles and methodologies of scientific inquiry and technological design; the distribution, use, properties, quality and stewardship of water systems; the properties of pure substances, the recognition or measurement of chemical changes in matter, and impacts of chemicals on humans; processes that affect biological and geological evolution, and how technologies can be used to monitor changes over time; the structures, functions, and processes of an animal cell, and the variety of single-celled organisms; relationships between micro-organisms and disease, human impacts on disease control, and the applications of biotechnology.

Achievement Level IV

Students performing at this level consistently demonstrate content mastery clearly beyond that required to be proficient and are well prepared for the next grade level.

Students demonstrate a considerable understanding of the application of principles and methodologies of scientific inquiry and technological design; relationships and unifying concepts of the distribution, use, properties, quality and stewardship of water systems; the application of the properties of pure substances, the recognition, measurement, and prediction of chemical changes in matter, and impacts of chemicals on humans; relationships and unifying concepts of the processes that affect biological and geological evolution, and how technologies can be used to monitor and predict changes over time; cellular structures, functions, and processes and explain how all three are interrelated and the variety of single-celled organisms; relationships between micro-organisms and disease, human impacts on disease control, biotechnology and apply them to real-world situations.