

The North Carolina English Language Arts Tests
Edition 3

Technical Report

Citable Draft

English I

March 2008

Prepared by:

Mildred Bazemore, Section Chief, Test Development
North Carolina Department of Public Instruction (NCDPI)

Melinda Taylor, Psychometrician, Test Development
North Carolina Department of Public Instruction (NCDPI)

Tammy Howard, Consultant, Test Development
North Carolina Department of Public Instruction (NCDPI)

Robert Brown, Statistician
North Carolina State University Technical Outreach for Public Schools (TOPS)

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

Inquiries or complaints regarding discrimination issues should be directed to:
J.B. Buxton, Deputy State Superintendent
6301 Mail Service Center :: Raleigh, NC 27699-6301
Telephone 919-807-3441 :: Fax 919-807-3445

Table of Contents

Chapter One: Introduction	10
1.1 Universal Participation	10
1.2 The North Carolina Testing Program	11
1.3 The North Carolina English I End-of-Course Test.....	13
Chapter Two: Test Development Process	15
2.1 Test Development Process for the North Carolina Testing Program	15
2.2 The Curriculum Connection	17
2.3 Test Specifications	17
2.4 Item Development	18
2.5 Item Format.....	19
2.6 Selection and Training of Item Writers	19
2.7 Reviewing Items for Field-Testing	20
2.8 Assembling Field Test Forms	21
2.9 Sampling Procedures and Field Test Sample Characteristics	22
2.10 Item Analysis	24
2.11 Classical Measurement Theory Analysis	24
2.12 Item Response Theory (IRT) Analysis	24
2.13 Differential Item Functioning Analysis	26
2.14 Expert Review	27
2.15 Criteria for Inclusion in Item Pool	28
2.16 Item Pool Parameter Estimates	29
2.17 Operational Test Construction	31
2.18 Setting the Target p-value for Operational Tests	31

2.19 Comparison of Item Pool p-Values with Operational p-Values	31
2.20 Review of Assembled Operational Tests	32
2.21 Setting the Test Administration Time	32
Chapter Three: Test Administration	34
3.1 Test Administration	34
3.2 Training for Test Administrators	34
3.3 Preparation for Test Administration	34
3.4 Test Security and Handling Materials	35
3.5 Student Participation	35
3.6 Alternate and Alternative Assessments.....	36
3.7 Testing Accommodations	36
3.8 Students with Limited English Proficiency	37
3.9 Medical Exclusions	37
3.10 Reporting Student Scores	37
3.11 Confidentiality of Student Test Scores	38
Chapter Four: Scaling, Equating, and Standard-Setting for the North Carolina EOC Test of English/Language Arts.....	39
4.1 Conversion of Raw Test Scores	39
4.2 Setting the Standards.....	39
4.3 Score Reporting for the North Carolina Tests	41
4.4 Achievement Level Descriptors	41
4.5 Achievement Level Cut Scores	42
4.6 Achievement Level Trends	43
4.7 Percentile Ranking	43
Chapter Five: Reports	44

5.1 Use of Test Score Reports Provided by the North Carolina Testing Program	44
5.2 Reporting by Student	44
5.3 Reporting by School	44
5.4 Reporting by the State	45
Chapter Six: Descriptive Statistics and Reliability	46
6.1 Descriptive Statistics for the First Operational Administration of the Test	46
6.2 Means and Standard Deviations for the First Operational Administration	46
6.3 Population Demographics for the First Operational Administration	46
6.4 Scale Score Frequency Distributions	46
6.5 Reliability of the North Carolina End-of-Course Test of English I.....	47
6.6 Internal Consistency of the North Carolina End-of-Course Test of English I.....	48
6.7 Standard Error of Measurement	48
6.8 Equivalency of Test Forms.....	50
Chapter Seven: Evidence of Validity	52
7.1 Evidence of Validity	52
7.2 Content Validity	52
7.3 Criterion-Related Validity	53
Chapter Eight: Quality Control Procedures.....	55
8.1 Quality Control Prior to Test Administration	55
8.2 Quality Control in Data Preparation and Test Administration	55
8.3 Quality Control in Data Input	56
8.4 Quality Control of Test Scores	56
8.5 Quality Control in Reporting	56
Glossary of Key Terms	57

References 60

Appendix A: Testing Code of Ethics.....62

Appendix B: Test Specifications English I Edition 3..... 66

Appendix C: Item Development Guidelines 73

Appendix D: Test Information Sheet for English I..... 75

Appendix E: Raw-to-Scale Score Conversion Tables.....87

Appendix F: End-of-Course Standard-Setting Report Executive Summary..... 92

Appendix G: Frequency Distribution Table for English I Scale Scores..... 98

List of Tables

Table 1:	Number of items field-tested for the North Carolina End-of-Course Test of English I.....	21
Table 2:	Field test population (2005–06) for the North Carolina End-of-Course Test of English I.....	23
Table 3:	Average item pool parameter estimates from stand-alone field tests for the North Carolina End-of-Course Test of English I.....	30
Table 4:	Comparison of p-value of item pool with p-values of assembled forms averaged across forms.....	31
Table 5:	Number of items and time allotted.....	33
Table 6:	Available assessments in the North Carolina Testing Program.....	36
Table 7:	Percent of students assigned to each achievement level by teachers.....	39
Table 8:	Achievement level descriptors for the North Carolina End-of-Course Test of English I.....	41
Table 9:	North Carolina End-of-Course Test of English I achievement levels and corresponding scale scores.....	42
Table 10:	Achievement level trends for the North Carolina End-of-Course Test of English I.....	43
Table 11:	Average scale score for the 2006–07 administration of the North Carolina End-of-Course Test of English I.....	46
Table 12:	Population demographics for the 2006–07 administration of the North Carolina End-of-Course Test of English I.....	46
Table 13:	Reliability indices averaged across the North Carolina End-of-Course Test of English I forms.....	48
Table 14:	Reliability indices averaged across the North Carolina End-of-Course Test of English I (Gender).....	48
Table 15:	Reliability indices averaged across the North Carolina End-of-Course Test of English I forms (Ethnicity).....	48
Table 16:	Reliability indices averaged across the North Carolina End-of-Course Test of English I forms (Other Characteristics).....	48
Table 17:	Range of standard error of measurement for the North Carolina End-of-Course Test of English I scale scores.....	49

Table 18: Instructional validity of the content of the North Carolina End-of-Course Test of English I 53

Table 19: Pearson correlation coefficient table for variables used to establish criterion-related validity for the North Carolina End-of-Course Test of English I 54

List of Figures

Figure 1: Flow chart of the test development process used in development of North Carolina tests 16

Figure 2: Thinking skills framework used to develop the North Carolina End-of-Course Test of English I (adapted from Marzano et al., 1988)..... 19

Figure 3: Typical item characteristic curve (ICC) for a 4-option multiple-choice item. 25

Figure 4: North Carolina End-of-Course Test of English I scale score frequency distribution..... 47

Figure 5: Standard errors of measurement on the North Carolina End-of-Course Test of English I forms50

Figure 6: Test characteristic curves for the North Carolina End-of-Course Test of English I forms51

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. Student’s 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. No Child Left Behind (NCLB) 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 mathematics once in grades 10–12; 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, Appendix).*

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.*

The North Carolina Testing Program includes multiple-choice assessments of reading comprehension and mathematics in grades 3 through 8 and English I, Algebra I, Geometry, and Algebra II. The program includes science assessments in grades 5 and 8, Biology, Physical Science, Chemistry, and Physics. In addition to the reading comprehension and mathematics tests, the North Carolina Testing Program includes social studies end-of course tests (Civics and Economics and U.S. History); writing assessments in grades 4, 7, and 10; the North Carolina Tests of Computer Skills; the North Carolina Competency Tests; and alternate assessments developed to validly measure student abilities in populations who are not able to access the general assessments even with accommodations.

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, 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.

The End-of-Grade tests in grades 3 through 8 in mathematics and grades 3 through 8 in reading comprehension are used for determining AYP at the elementary and middle school levels. At the high school level, the End-of-Course tests in English I, Algebra I, and the grade 10 Writing assessment are used for determining AYP. For students who are not able to access the general assessments, the corresponding alternate or alternative assessment is used.

In 2006, the North Carolina State Board of Education approved new graduation standards. These standards require that

Effective with the class entering ninth grade for the first time in the 2006–2007 school year, students who are following the career preparation, college technical preparation, or college/university preparation courses of study shall meet the following exit standards:

- (A) successfully complete a senior project that is developed, monitored, and scored within the LEA using state-adopted rubrics and*
- (B) score at proficiency level III or above on the end-of-course assessment for English I, U.S. History, Biology, Civics and Economics, and Algebra I.*

(16 NCAC 6D .0503 State graduation requirements, section E subsection 2).

The Grade-Level Proficiency Guidelines, approved by the State Board of Education (February, 1995), established Level III (of those achievement levels listed above) as the standard for each grade level. The EOC tests measure a student's mastery of course-level material.

Scale scores are derived from a raw score, the number correct score, for the test. Each test has a translation table that provides a scale score for each raw test score. Scale scores are reported with four achievement levels, which are predetermined academic achievement standards. The descriptors below are the State Board of Education adopted generic achievement level descriptors.

Level I: Students performing at this level do not have sufficient mastery of knowledge and skills in a particular subject area to be successful at the next grade level.

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

Level III: Students performing at this level consistently demonstrate mastery of the grade-level subject matter and skills and are well prepared for the next grade.

Level IV: Students performing at this level consistently perform in a superior manner clearly beyond that required to be proficient at grade level.

The content-specific performance-level descriptors for English I are provided for each assessment (pp. 41–42).

1.3 The North Carolina English I End-of-Course Test

This technical report for the third edition of the North Carolina English I End-of-Course Test discusses tests aligned with the North Carolina English Language Arts (ELA) 2004 *Standard Course of Study (SCS)*. Following a five-year revision cycle, the North Carolina State Board of Education adopted the ELA *SCS* in March 2004 to replace the 1999 *SCS*. The English I EOC test was field-tested during the fall 2005 and spring 2006, and the first operational year was 2006–07.

The purpose of this document is to provide an overview and technical documentation for the North Carolina English I End-of-Course Test, 3rd Edition. Chapter One provides an overview of the English I test. Chapter Two describes the test development process. Chapter Three outlines the test administration. Chapter Four describes the construction of the developmental scale, 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 first operational year, reliability indices, and evidence of validity. Chapter Eight is an overview of quality control procedures.

Chapter Two: Test Development Process

2.1 Test Development Process for the North Carolina Testing Program

In June 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 flow chart of those events is found in Figure 1.

Figure 1: Flow chart 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.

2.2 The Curriculum Connection

North Carolina wants its students to graduate with the skills necessary to compete in the global marketplace, to be prepared for further education, and to participate effectively as citizens.

The previous revision to the ELA North Carolina *Standard Course of Study* (NCSCS) was 1999. Following the North Carolina five-year revision cycle, the 2004 revisions are “based upon surveys of the effectiveness of the previous curriculum, current educational research, and input from public school teachers, administrators, college and university faculties, parents, and business and community leaders. In addition, the revised curriculum is based on national curriculum standards developed jointly by the National Council of Teachers of English and the International Reading Association and on language arts standards developed by the National Assessment of Educational Progress (SCS, p. 3). According to the SCS, “The ultimate purpose of the English Language Arts curriculum is to teach students the language abilities they need to communicate effectively as individuals and as contributing members of society (p.7). The NCSCS is available at <http://www.ncpublicschools.org/curriculum/languagearts/scos/2004/27english1>.

The North Carolina ELA SCS clearly defines a curriculum focused on what students will need to know and be able to do to be successful and contributing citizens in our state and nation in the years ahead. As defined in the 2004 North Carolina ELA SCS, the goals of the curriculum are for students to:

- Express reflections and reactions to literature and to personal experience;
- Explain meaning, describe processes, and answer research questions;
- Evaluate communication and critique texts;
- Make and support an informed opinion;
- Participate in conversations about and written analysis of literary genres, elements, and traditions; and
- Use knowledge of language and standard grammatical conventions.

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, item tryout, and item 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 English I End-of-Course Test is legislated by General Statute *115C-174.10* and focuses on the measurement of individual student ELA skills and knowledge as outlined in the NCSCS.

Test specifications for the North Carolina ELA 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 A. 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 into level of difficulty;
- (2) Percentage of test questions that measure a specific goal or objective; and
- (3) Number of selections from each of the genres cited in the NCSCS.

Test blueprints, specific layouts designed to ensure the parallel construction of multiple test forms, 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 super-ordinate 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.

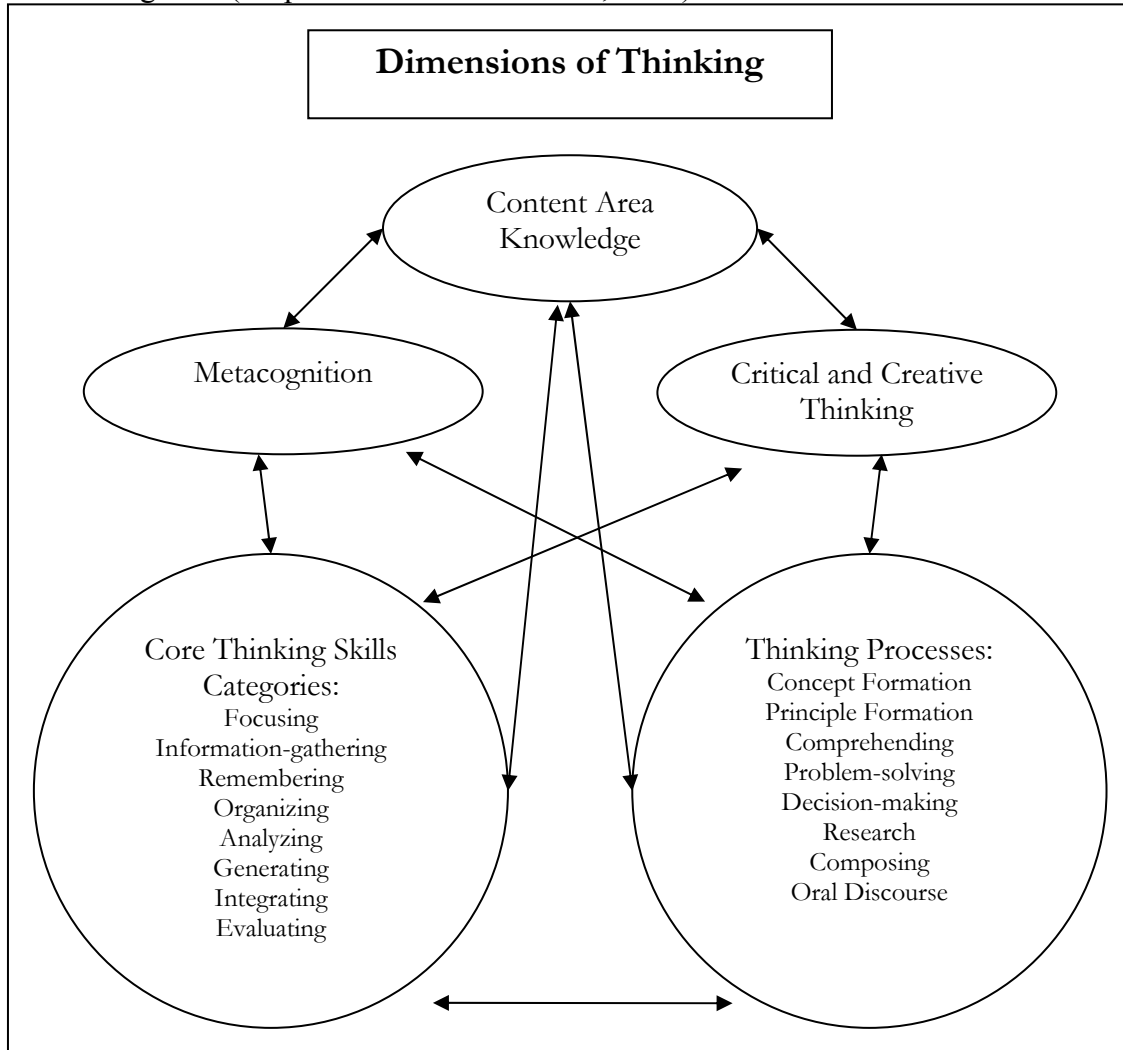
2.4 Item Development

Each item is written to be aligned with a specific objective in the NCSCS. Items on the English I test are developed using level of difficulty and Marzano's thinking skill level (Marzano et al., 1988). Item writers use these frameworks when developing items. The purpose of the categories is to ensure a balance of items across difficulty as well as a balance of items across the different cognitive levels among the items on the English I test.

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. Easy items are those items that the item writer believes can be answered correctly by approximately 70% of the examinees. Medium items can be answered correctly by 50–60% of the examinees. Difficult items can be answered correctly by approximately 20–30% 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 the North Carolina End-of-Course Test of English I (adapted from Marzano et al., 1988)



2.5 Item Format

Items on the English I tests are four-foil, multiple-choice items. Standards for item construction, such as concisely written stems, plausible distractors, and parallel distractors are adhered to in the development of items.

2.6 Selection and Training of Item Writers

Once the test blueprints were finalized from the test specifications for the revised editions of the North Carolina English I End-of-Course Test, North Carolina educators were recruited and trained to write new items for the tests. The diversity among the item writers and their knowledge of the current NCSCS were addressed during recruitment. The use of North Carolina educators to develop items strengthened the instructional validity of the items. Some items were developed through an external vendor; however,

the vendor was encouraged to use North Carolina educators in addition to professional item writers to generate items that would align with the NCSCS for English I. Specifically, the contract required that at least half of the contractor-supplied items would be written by North Carolina educators.

Potential item writers received training and materials designed in accordance with the English I 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 B. 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.

2.7 Reviewing Items for Field-testing

To ensure that an item was developed to NCSCS standards, each item went through a detailed review process prior to being placed on a field test. This review is represented by Step 9 on the Test Development Flow Chart (Figure 1). A new group of North Carolina educators was recruited to review items. Once items had been through an educator review, test development staff and curriculum specialists reviewed each item. Items were also reviewed by educators and/or staff familiar with the needs of students with disabilities and limited English proficiency.

The criteria for evaluating each written item included the following:

1) Conceptual

- objective match (curricular appropriateness)
- thinking skill match
- fair representation
- lack of bias/sensitivity
- clear statement
- one best answer
- credible foils
- technical correctness

2) Language

- appropriate for age
- correct punctuation
- spelling and grammar
- lack of excess words
- no stem or foil clues
- no negative in foils (unless it fits the objective)

3) Format

- logical order of foils
- familiar presentation style, print size, and type

- correct mechanics and appearance
- equal/balanced length foils

4) Art/Graphics

- necessary
- clean
- relevant
- unbiased

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

2.8 Assembling Field Test Forms

Prior to creating an operational test, items for each grade level or course area were assembled into field test forms. Field test forms were organized according to the blueprints for the operational tests. All forms were administered as stand-alone field tests. All items were aligned with the 2004 North Carolina *Standard Course of Study (SCS)* content standards. Prior to field test administration, North Carolina educators reviewed the assembled field test forms for clarity, correctness, potential bias or sensitivity, cuing of items, and curricular appropriateness, following steps similar to operational test review.

The initial round of field tests for the Edition 3 English I tests was stand-alone, rather than embedded field tests (See Table 1). The 2004 *SCS* for grades K–12 was first implemented instructionally in academic year 2005–06, and the stand-alone field test for English I was administered in fall 2005 and spring 2006. The fall administration was for schools on a block, or semester, schedule where instruction in English I was completed by December or January in the school year 2005–06. The spring administration was for schools also on the block schedule and for students on a traditional schedule where instruction was completed at the end of the school year 2005–06.

This transition between curricula meant that in the field test year, students were actually receiving instruction in both the old and new curricula. Thus, in addition to the field test forms aligned to the 2004 *SCS*, operational forms aligned to the 1999 *SCS* were included in the field test spiral for research purposes. Data from those forms, other than in the context of the research (see Chapter 4) is not included in this technical report.

Table 1: Number of items Field-tested for the North Carolina End-of-Course Test of English I

Administration(s)	Number of Forms	Number of Items per Form	Total Number of Items
Fall 2005/Spring 2006	10	72 to 78	751

2.9 Sampling Procedures and Field Test Sample Characteristics

Sampling for stand-alone field-testing of the North Carolina Tests is typically accomplished using stratified random sampling of schools with the goal being a selection of students that is representative of the entire student population in North Carolina. Table 2 shows the demographic characteristics of the sample for the stand-alone field tests of the third edition English I test. Stratifying variables include

- gender,
- ethnicity,
- region of the state,
- free/reduced lunch,
- students with disabilities,
- students with limited English proficiency, and
- previous year's test scores.

Beginning with the first operational version of the third edition English I test, field test items are embedded within each form to supplement the item pools. Embedded field test items are grouped by selections. The experimental selections are placed in operational forms, and the operational forms are spiraled within a classroom to obtain a randomly equivalent group of examinees on each form. This results in a demographic distribution nearly identical to that of the full population.

Table 2: Field test population (2005–06) for the North Carolina End-of-Course Test of English I

Course	N	Gender		Ethnicity						English Language Proficiency Status
		% Male	% Female	% Asian	% Black	% Hispanic	% American Indian	% Multiracial	% White	% LEP (Limited English Proficiency)
English I	16,382	50.2	49.8	2.9	29.3	6.8	1.2	2.6	57.2	3.1

Notes: Fall 2005 and spring 2006 stand-alone field-testing for English I was accomplished by selecting demographically representative samples. The N value listed in this table represents examinees taking field test forms aligned with the 2004 North Carolina *Standard Course of Study (SCS)*. Additional examinees took test forms containing 1999-SCS-aligned content administered for research purposes. The percentages for demographic categories are for all examinees with available demographic data, including those examinees taking a research form.

2.10 Item Analysis

Field-testing provides important data for determining whether an item will be retained in the operational item pool. The North Carolina Testing Program uses both classical measurement theory and item response theory (IRT) 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.

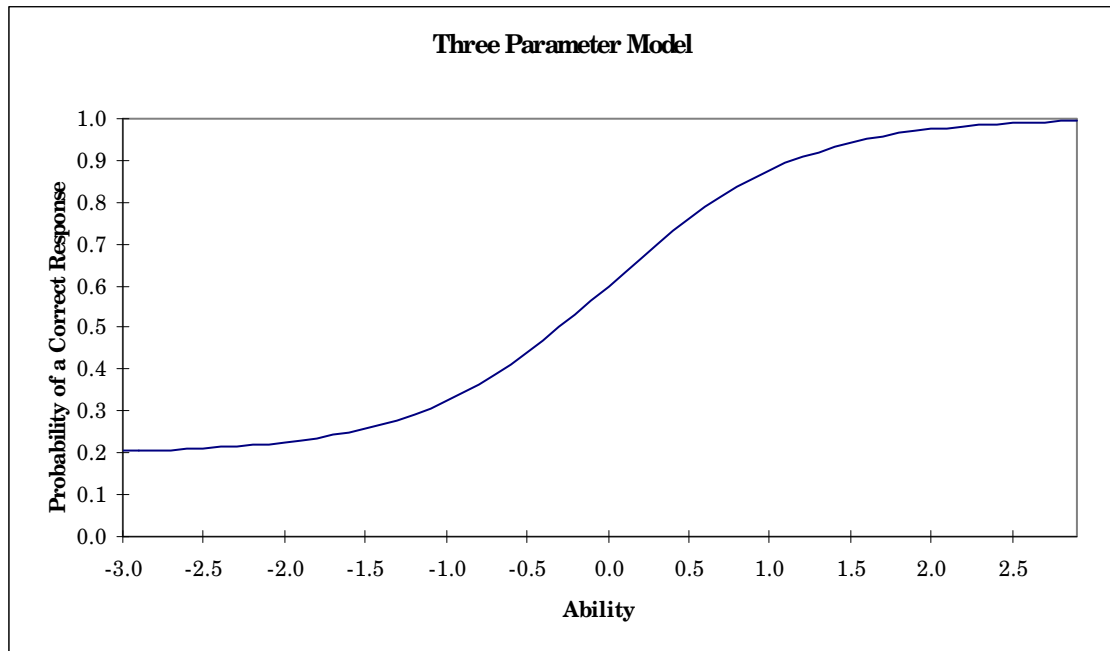
Field test data were analyzed by the North Carolina Department of Public Instruction (NCDPI) psychometric staff. Item statistics and descriptive information were then included on the item record for each item. The item records contained the statistical, descriptive, and historical information for an item, a copy of the item as it was field-tested, comments by reviewers, and curricular and psychometric notations.

2.11 Classical Measurement Theory Analysis

For each item, the p-value (proportion of examinees answering an item correctly), the standard deviation of the p-value, and the point-biserial correlation between the item score and the total test score were computed using SAS. In addition, frequency distributions of the response choices were tabulated. While the p-value is an important statistic and one component used in determining the selection of an item, the North Carolina Testing Program also uses IRT to provide additional item parameters to determine the psychometric properties of the North Carolina English/Language Arts tests.

2.12 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). IRT meets these needs and provides two additional advantages: the *invariance of item parameters* and the *invariance of ability parameters*. 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. The following figure shows the ICC for a typical 4-option multiple-choice item.

Figure 3: Typical item characteristic curve (ICC) for a 4-option multiple-choice item

The three-parameter logistic model (3PL) of IRT, the model used in generating the item statistics, 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 3PL 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 3PL model is

$$P_i(\theta) = c_i + \frac{1 - c_i}{1 + \exp[-Da_i(\theta - b_i)]}$$

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 (the slope of a typical item is 1.00)
- b —the threshold, “difficulty parameter,” the point on the ability scale where the probability of a correct response is 50% **when $c = 0$** (the threshold of a typical item is 0.00)
- c —the asymptote, “pseudo-guessing parameter,” the proportion of the examinees who got the item correct but did poorly on the overall test (the theoretical asymptote of a typical 4-choice item is 0.25)
- 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 [$a \sim \text{lognormal}(0, 0.5)$, $b \sim N(0,2)$, and $c \sim \text{Beta}(6,16)$].

2.13 Differential Item Functioning Analysis

It is important to know the extent to which an item on a test performs differently for different students. As a third component of the item analysis, differential item functioning (DIF) analyses examine the relationship between the score on an item and group membership, while controlling for ability, to determine if an item may be biased toward a particular gender or ethnic group. In developing the North Carolina English/Language Arts tests, the North Carolina Testing Program staff used the Mantel-Haenszel procedure to examine 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 & Holland, 1993; Camilli & Shepherd, 1994). For example, females might serve as the focal group and males might serve as the reference group to determine if an item may be biased toward or against females.

The Mantel-Haenszel (MH) chi-square statistic (only used for 2×2 tables) tests the alternative hypothesis that a linear association exists between the row variable (score on the item) and the column variable (group membership). The X^2 distribution has one degree of freedom (df) and its significance is determined by the correlation between the

row variable and the column variable (SAS Institute, 1985). The MH Log Odds Ratio statistic in SAS was used to determine the direction of DIF. This measure was obtained by combining the odds ratios (a_j) across levels with the formula for weighted averages (Camilli & Shepherd, 1994, p. 110).

For the Mantel-Haenszel statistic, the null hypothesis is that there is no relationship between score and group membership: the odds of getting the item correct are equal for the two groups. The null hypothesis was not rejected when the odds ratio equaled 1. For odds ratios greater than 1, the interpretation was that an individual at score level j of the Reference Group had 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 was that an individual at score level j of the Focal Group had a greater chance of answering the item correctly than an individual at score level j of the Reference Group. The Breslow-Day Test was used to test whether the odds ratios from the j levels of the score were all equal. When the null hypothesis was true, the statistic was distributed approximately as a chi-square with $j-1$ degrees of freedom (SAS Institute, 1985).

The ethnic (Black/White) and gender (Male/Female) bias flags were determined by examining the significance levels of items from several forms and identifying a typical point on the continuum of odds ratios that was statistically significant at the $\alpha = 0.05$ level.

2.14 Expert Review

All items, statistics, and comments were reviewed by curriculum specialists and testing consultants. Items found to be inappropriate for curricular or psychometric reasons were deleted. In addition, items flagged for exhibiting ethnic or gender DIF were then reviewed by a bias review committee. Differential item functioning is a purely statistical judgment without regard to the actual content of the item; the determination of actual bias is a qualitative judgment based on the content of the item.

The bias review committee members, selected because of their knowledge of the curriculum area and their diversity, evaluated test items with a DIF flag using the following questions:

1. Does the item contain language that is not commonly used statewide or has different connotations in different parts of the state or in different cultural or gender groups?
2. Does the item contain any local references that are not a part of the statewide curriculum?
3. Does the item portray anyone in a stereotypical manner? (These could include activities, occupations, or emotions.)
4. Does the item contain any demeaning or offensive materials?
5. Does the item have offensive, stereotyping, derogatory, or proselytizing religious references?

6. Does the item assume that all students come from the same socioeconomic background? (e.g., a suburban home with two-car garage)
7. Does the artwork adequately reflect the diversity of the student population?
8. Are there other bias or sensitivity concerns?

An answer of yes to any of these questions resulted in the unique item production number being recorded on an item bias sheet along with the nature of the bias or sensitivity. Items that were consistently identified as exhibiting bias or sensitivity were flagged for further review by curriculum specialists.

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

2.15 Criteria for Inclusion in Item Pool

All of the item parameter data generated from the above analyses were used to determine if an item displayed sound psychometric properties. Items could be potentially be flagged as exhibiting psychometric problems or bias due to ethnicity/race or gender according to the following criteria:

Items with these characteristics were deleted:

- weak discrimination—the slope (a parameter) was less than 0.50
- low correlation with total score—the item correlation (r-biserial) was less than 0.15
- guessing—the asymptote (c parameter) was greater than 0.45
- too difficult—the threshold (b parameter) was greater than 3.0 or the p-value was less than 0.10

Items with these characteristics were used sparingly (held in reserve):

- weak discrimination—the slope (a parameter) was between 0.70 and 0.50
- low correlation with total score—the item correlation (r-biserial) was between 0.25 and 0.15
- guessing—the asymptote (c parameter) was between 0.35 and 0.45
- too difficult—the threshold (b parameter) was between 2.5 and 3.0; or the p-value between 0.10 and 0.15
- too easy—the threshold (b parameter) was between -2.5 and -3.0 ; or the p-value was between 0.85 and 0.90.

Items with these characteristics underwent additional reviews:

- ethnic bias—the log odds ratio was greater than 1.50 (favored whites) or less than 0.67 (favored blacks)
- gender bias—the log odds ratio was greater than 1.50 (favored

females) or less than 0.67 (favored males)

Items with a threshold less than -3.0 or p-value greater than 0.90, provided all other statistical and content information supported keeping the item, were submitted for consideration in an alternative assessment targeted toward students with persistent academic disabilities.

The average item pool parameter estimates based on field test data are provided below.

2.16 Item Pool Parameter Estimates

See Table 3 on the following page.

Table 3: Average item pool parameter estimates from stand-alone field tests for the North Carolina End-of-Course Test of English I

Grade or Course	Biserial Correlation	IRT Parameters			P-value	DIF (Odds Ratio Logit)	
		Threshold (b)	Slope (a)	Asymptote (c)		Ethnicity	Gender
English I	.541	.210	1.152	.215	.5737	1.112	.987

Note: The item pool averages shown in this table are for all items from the field-testing that upon post-field test review of content and psychometric properties were retained as primary candidates for potential use on operational test forms.

2.17 Operational Test Construction

The final item pool was based on approval by content and curriculum experts for curricular match and testing experts and psychometricians for psychometrically sound item performance. Once the final items were identified for the item pool, operational tests were constructed according to the test blueprints. For a summary of the test specifications, see Appendix A. Six forms were developed for the English I test.

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 answer an item correctly. So an item with a p-value of 0.75 was correctly answered by 75% of the students who answered the item during the field test, and one might expect that roughly 75 of 100 examinees will answer it correctly when the item is put on an operational test. An easy item has a p-value that is high, which 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 answered the item correctly during field-testing. Note that items usually have higher p-values on the operational form than on stand-alone field tests, due to factors that may include higher motivation on the operational test, which has stakes for the student and/or improved instruction in the content in the second and subsequent years of a new curriculum.

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 would be 0.625, which is the theoretical average of a student getting 100% correct on the test and a student scoring a chance performance (25% for a 4-foil multiple-choice test). That is $(100 + 25)/2$.

The actual target was 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 0.625 as possible, it is often the case that the target p-value is set between the ideal 0.625 and the average p-value of the item pool. The average p-value and the target p-value for operational forms are provided below for comparison.

2.19 Comparison of Item Pool p-Values with Operational p-Values

Table 4: Comparison of p-value of item pool with p-values of assembled forms averaged across forms

Course	Ave. p-Value of Item Pool	Ave. p-Value of Assembled Forms
English I	0.5737	0.576

To develop equivalent forms, the test forms were balanced on P+, the sum of the p-values of the items. To the extent possible, the sections were balanced on slope. Although all form-level values are reported as an average across forms, actual P+ differences between forms within the same grade were less than 0.01.

The first operational administration for the new 2004-SCS-aligned End-of-Course English 1 test forms comprised two primary testing windows, Fall 2006 and Spring 2007. These tests were administered with experimental material embedded. The table below shows the average p-values for the operational items and for the experimental items. Note that, principally due to increased examinee motivation, the p-values for the operational items were higher at this administration than the p-values for the same items at the 2005-06 stand-alone field-testing (see table 4 in this chapter.) The experimental items were used for subsequent test development.

Table 4b: 2006-07 Operational Test Administrations: End-of-Course English I

	Operational Average P-Value			Experimental Average P-Value		
	Comp	Text Analysis	Overall	Comp	Text Analysis	Overall
English I	.6923	.6224	.6473	.6359	.5696	.5917

Table 4c: Fall 2006 Operational Test Administration: End-of-Course English I

	Operational Average P-Value		
	Composition	Textual Analysis	Overall
English I	.6868	.6111	.6382

2.20 Review of Assembled Operational Tests

Once forms were assembled to meet test specifications, target p-values, and item parameter targets, a group of North Carolina educators and curriculum supervisors then reviewed the assembled forms. Each group of subject-area teachers and curriculum supervisors worked independently of the test developers. 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 (curricular validity);
- the content of test forms should reflect the goals and objectives as 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 (free from test/item bias); and
- an item should have one and only one best answer that is right; 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 as well as recording their responses on a separate sheet) and to provide comments and feedback next to each item. After reviewing all the forms, 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 test review ratings along with the comments were aggregated for review by NCDPI curriculum specialists and testing consultants. Items determined to be problematic at this point were replaced, and the forms rebalanced. Items may have been removed from a form because of cuing, overemphasis on a particular subtopic (e.g., several items on main idea or purpose), or for maintaining statistical equivalency. If a form has more than 10% of its items replaced as a result of this process, the NCDPI psychometric policy is to send the form through review again, as it is no longer really the same form that was reviewed previously. No test forms exceeded this criterion. As a final review, test development staff members with input from curriculum staff, content experts, and editors conducted a final psychometric, content, and grammar check for each test form.

2.21 Setting the Test Administration Time

Ultimately, NC tests are designed to be untimed; however, there is the need to provide information to LEA and district personnel for scheduling purposes. Timing is an important consideration in the construction of NC tests. The timing of field test administrations is routinely analyzed to aid in determining the appropriate timing of the operational administration. Tests in NC are designed such that the length allows at least 95% of students taking the test to complete all items. The total number of items and approximate testing time (in minutes) for the North Carolina End-of-Course English I test is provided below.

Table 5: Number of items and time allotted

Course	Number of Items	Time Allotted (includes short breaks and general instructions)
English I	80	130

Chapter Three: Test Administration

3.1 Test Administration

The North Carolina End-of-Course Test of English I is a multiple-choice test administered to all students enrolled in English I. All end-of-course tests are administered within the final ten days of the course to students enrolled for credit in courses where end-of-course tests are required. The purpose of end-of-course tests is to sample a student's knowledge of subject-related concepts specified in the North Carolina *Standard Course of Study* and to provide a global estimate of the student's mastery of the material in a particular content area. The English I end-of-course test was developed to provide accurate measurement of individual student knowledge and skills specified in the English I component of the North Carolina *Standard Course of Study*.

The English I end-of-course test assesses the goals for English I described in the North Carolina *Standard Course of Study*. The test consists of two separate sections: composition and textual analysis. The composition section, part 1, contains four student draft papers with five to eight associated questions for each, for a total of 28 questions. The student draft papers represent expressive, informational, argumentative, or critical text modes of writing. Students are asked to read the student drafts and respond to related questions about editing and composition.

The textual analysis section, part 2, contains seven literary selections and 52 questions. The selections include poetry; informational, fictional, or expressive nonfiction texts; and either an argumentative or critical text. Each selection is followed by six to nine associated questions. The students are asked to answer related questions in which they must analyze the text for general comprehension as well as author's craft and strategies. Additional information is provided in the Test Information Sheet for English I (Appendix C), which is posted on the department's Web site for access by teachers, parents, and students.

3.2 Training for Test 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 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), accommodated test administrations, test security (storing, inventorying, and returning test materials), and the *Testing Code of Ethics*.

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 must thoroughly read the *Test Administrator's*

Manual and the codified North Carolina *Testing Code of Ethics* prior to actual test administration. Test administrators must also follow the instructions given in the *Test Administrator's Manual* to ensure a standardized administration and must read aloud all directions and information to students as indicated in the manual.

3.4 Test Security and Handling Materials

Compromised secure tests result in invalid test scores. To avoid 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 school 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).

All students, including students with disabilities, enrolled in a yearlong (i.e., traditional calendar) course for credit must be administered the end-of-course test, which may be a corresponding alternate or alternative assessment if so indicated by the student's IEP, during the final ten days of the course. In schools that operate on a block or semester schedule, all students, including students with disabilities, who are enrolled in a course for credit must be administered the EOC test during the final five days of the course. Students enrolled for credit in a course that has an end-of-course test must be administered the EOC test. Students who are repeating the course for credit must also be administered the EOC test. The student's most recent test score will be used for the purpose of state accountability. In addition, starting with the 2001–2002 school year, LEAs shall use results from all multiple-choice EOC tests (including English I) as at least twenty-five percent of the student's final grade for each respective course. LEAs shall adopt policies regarding the use of EOC test results in assigning final grades.

In 2006, the NC State Board of Education revised policy HSP-N-004 (16 NCAC 6D.0503): students entering the ninth grade for the first time in 2006–07 and beyond are now required to perform at Achievement Level III (with one standard error of measurement) or above on five required EOC

assessments in order to graduate. Multiple retest opportunities are available; however, the first test score is used for the purpose of AYP and federal accountability

3.6 Alternate and Alternative Assessments

To provide access for students with disabilities and students with limited English proficiency, the North Carolina Testing Program makes available the North Carolina Checklist of Academic Skills (NCCLAS) for English I.

Table 6: Available assessments in the North Carolina Testing Program

Subject	General		Modified Format	Modified Achievement Standards	Alternate Achievement Standards
	Unaccommodated	with Accommodations	NCCLAS	<i>NCEXTEND2</i>	<i>NCEXTEND1</i>
English I	X	X	X		

The NCCLAS is an assessment process in which teachers utilize a checklist to evaluate student performance on curriculum benchmarks in the area of reading. Student performance data are provided to the NCDPI at the end of the school year (summative), although teachers gather evidence throughout the year. The NCCLAS measures competencies on the North Carolina *Standard Course of Study*. The Individualized Education Program (IEP) team determines if a student, due to the nature of his/her special needs, is eligible to participate in the NCCLAS. Typically, students who are being assessed on the NCCLAS should be those students who are unable to access the paper and pencil test, even with accommodations. Additionally, students who are limited English proficient (that is, students who have been assessed on the state identified English language proficiency tests as below Intermediate High in reading and been enrolled in U.S. schools for less than two years) may also participate in NCCLAS for reading, mathematics, writing, and/or science. These students have received instruction on the grade-level academic content standards outlined in the NCSCS and are held to the same grade-level academic achievement standards.

3.7 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 and 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 <http://www.ncpublicschools.org/accountability/policies/tswd/>. 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.

3.8 Students with Limited English Proficiency

Per HSP-C-021(d), last revised in April 2007, students identified as limited English proficient shall be included in the statewide testing program as follows: standard test administration, standard test administration with accommodations, or the state-designated alternate assessment. Students identified as limited English proficient who have been assessed on the state English language proficiency tests as below Intermediate/High in reading and who have been enrolled in United States schools for less than two years may participate in the state designated alternate assessment in the areas of reading and mathematics at grades 3 through 8 and 10, science at grades 5 and 8, and in high school courses in which an end-of-course assessment is administered. To be identified as limited English proficient students must be assessed using the state English language proficiency tests at initial enrollment. All students identified as limited English proficient must be assessed using the state English language proficiency test annually thereafter during the spring testing window. A student who enrolls after January 1 does not have to be retested during the same school year.

Schools must administer state reading, mathematics, end-of-course assessments, and writing tests for students identified as limited English proficient who score at or above Intermediate/High on the state English language proficiency reading test during their first year in US schools. Results from these assessments shall be included in the ABCs and AYP. Additionally, schools must include students previously identified as limited English proficient, who have exited limited English proficient identification during the last two years, in the calculations for determining the status of the limited English proficient subgroup for AYP only if that subgroup already met the minimum number of 40 students required for a subgroup.

3.9 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.

3.10 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 district-wide 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 district-wide 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, the computer skills tests, and the specified end-of-course 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.

3.11 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, Equating, and Standard-Setting for the North Carolina EOC Test of English I

The North Carolina EOC Tests of English I scores are reported as scale scores, achievement levels, and percentiles. Scale scores are advantageous in reporting because:

- scale scores can be used to compare test results when there have been changes in the curriculum 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, and
- scale scores can be used to minimize differences among various forms of the tests.

4.1 Conversion of Raw Test Scores

Each student's score is determined by calculating the number of items he or she answered correctly and then converting the sum to a developmental scale score. 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 parameters (threshold, slope, and asymptote) for each item. The software implements the algorithm described by Thissen and Orlando (2001, pp. 119–130). Because different items are placed on each form of a subject's test, unique score conversion tables are produced for each form of a test for each grade or subject area. For example, EOC English I has six parallel test forms. Therefore, the scanning and reporting program developed and distributed by the NCDPI uses six scale-score conversion tables, one for each parallel form. Each scale score has an associated conditional standard error of measurement. The raw-to-scale conversion tables are provided as Appendix D.

4.2 Setting the Standards

For tests developed in the North Carolina Testing Program, academic achievement standard-setting, the process of determining cut scores for the different achievement levels, has historically been accomplished through the use of contrasting groups, and this method continues to be one source of information that is considered when setting standards. Contrasting groups is an examinee-based method of standard-setting that involves the categorization of students into the four achievement levels by expert judges who are knowledgeable of students' achievement in various domains outside of the testing situation. These judgments are then compared to students' actual score distribution. For the English I end-of-course test, North Carolina teachers were considered as expert judges under the rationale that teachers were able to make informed judgments about students' academic achievement because they had observed the breadth and depth of the students' work during the school year.

For the academic achievement standard-setting for the third edition of the NC EOC Test of English I, approximately 109,989 students were placed into categories. Teachers indicated the achievement level for each student who participated in both the fall and spring administrations of the operational test. Table 7 shows the percentage of students classified into each achievement level.

Table 7: Percent of students assigned to each achievement level by teachers

Course	Level I	Level II	Level III	Level IV
English I	8.07	22.55	46.81	22.56

For the third edition of the North Carolina EOC Test of English I, the proportions of students expected to score in each of the four achievement levels were collected in the fall of the first operational (pilot) administration (Fall 2006). It is important to note that these proportions were not significantly different when the entire year's data were collected. These proportions were applied to the distribution of student scores to arrive at one possible set of cut scores. Stratifying variables were discussed previously in Section 2.9, Sampling Procedures and Field Test Sample Characteristics. The sample served as the source for both the initial contrasting groups proportions and the score distribution to which the proportions were applied to obtain the interim cut scores. The interim cut scores were used for the entire 2006–2007 school year.

When applying the contrasting groups approach to standard-setting for the third edition of the North Carolina EOC Test of English I, the scale scores from the pilot year were displayed in a frequency table from lowest to highest scale score. Using the classifications for achievement level I as an example, 8.07% of 109,989 would be 8,876 scores. The cumulative frequency distribution allows one to find the scale score below which 8,876 students scored. This scale score is thus one possible cut-off between Level I and Level II. The process continued for each of the levels until all cut scores had been derived. Of course, with a small number of scale score points relative to the number of examinees, it was very unlikely that a score would have exactly 8,876 students at or below that score. The rule of thumb was to get as close as possible to the contrasting groups percentage without exceeding it. This results in giving students the advantage at Levels I, II, and III, and slightly exceeding the teacher-adjudicated proportion of examinees in level IV. It should be noted that to avoid a further inflation of children categorized as Level IV, the percentage of students not rated was removed from the cut score calculations.

Subsequent standard-setting procedures were conducted in the fall of 2007. The methodology used to obtain additional information was the item-mapping (or bookmarking) method, as described by Mitzel, Lewis, Patz, & Green (2001). In both cases, the item-mapping process was moderated by an outside vendor. A report from this standard-setting can be found in Appendix E.

Any standard-setting method is a reasoned judgment. Knowing that any one method might give a different set of cut scores, multiple ways of deriving cut scores are always examined in an attempt to arrive at the most appropriate set of cut scores.

The final standards were set using impact data from the first full year's operational data, looking at the information obtained from all standard-setting methods and considered in the context of the current assessment climate in the state. All of this information was summarized and shared with a group of stakeholders representing curriculum, accountability, and exceptional children. This final panel engaged in a standard-setting method described as "Reasoned Judgment." As described by Roeber (2002), a group such as an expert panel, a representative group of users, or a policymaker group examines the data and divides the full range of possible scores into the desired categories. The reasoned judgment methodology used for the North Carolina EOC Test of English I involved looking at the quantitatively derived cut scores as well as considering the policy and practice implications of the cut scores. The panels' recommendation was presented to the State Board of Education and adopted as policy in November 2007.

4.3 Score Reporting for the North Carolina Tests

Scores from the North Carolina End-of-Course Test of English I are reported as scale scores, achievement levels, and percentile ranks. The scale scores are computed through the use of raw-to-scale score conversion tables. The scale score determines the achievement level in which a student falls.

Score reports are generated at the local level to depict performance for individual students, classrooms, schools, and local education agencies. The 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, limited English proficient 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 the data for school accountability, student accountability (grades 3, 5, and 8), and to satisfy other federal requirements under the No Child Left Behind Act of 2001.

4.4 Achievement Level Descriptors

The four achievement levels in the North Carolina Testing Program are given below.

Table 8: Achievement level descriptors for the North Carolina End-of-Course Test of English I

Achievement Levels for the North Carolina Testing Program	
Level I	<p>Students performing at this level do not have sufficient mastery of knowledge and skills of the course to be successful at a more advanced level in the content area.</p> <p>Students performing at Achievement Level I demonstrate the need to develop the composition and reading comprehension skills required in the English I North Carolina <i>Standard Course of Study</i>. Students inconsistently identify and correct rudimentary language convention errors (such as incorrect verb usage, end punctuation errors, double negatives, capitalization errors, and non-standard verb forms). Students show little to no evidence of reading skills and strategies required to comprehend a variety of ninth-grade-level expressive, informational, argumentative, critical, and literary texts.</p>
Level II	<p>Students performing at this level demonstrate inconsistent mastery of knowledge and skills of the course and are minimally prepared to be successful at a more advanced level in the content area.</p> <p>Students performing at Achievement Level II demonstrate inconsistent application of the composition and reading skills required in the English I North Carolina <i>Standard Course of Study</i>. Students inconsistently apply knowledge of grammar and language usage to identify and correct language convention errors in spelling, punctuation, and simple sentence structure. Students show an initial understanding of basic literary devices and elements (such as symbolism, plot structure, and figurative language). Students may also inconsistently apply strategies such as determining the meaning of unfamiliar vocabulary through context clues and identifying the main idea and supporting details. Students show inconsistent literal comprehension of a variety of ninth-grade-level expressive, informational, argumentative, critical, and other literary works.</p>

Level III	<p>Students performing at this level consistently demonstrate mastery of the course subject matter and skills and are well prepared for a more advanced level in the content area.</p> <p>Students performing at Achievement Level III typically demonstrate composition and reading comprehension skills required by the English I North Carolina <i>Standard Course of Study</i>. Students typically demonstrate an understanding of conventional written expression by editing sentences for correctness, clarity, and style. Students can infer, generalize, draw conclusions, and make connections between texts. Students can analyze the impact of details and literary elements and devices (such as characterization, dialect, imagery, symbolism, main idea, purpose, context, and figurative language) on the work as a whole. Students are typically able to comprehend and analyze a variety of ninth-grade-level expressive, informational, argumentative, critical, and other literary texts.</p>
Level IV	<p>Students performing at this level consistently perform in a superior manner clearly beyond that required to be proficient in the course subject matter and skills and are very well prepared for a more advanced level in the content area.</p> <p>Students performing at Level IV demonstrate a strong command of the composition and reading comprehension skills required by the English I North Carolina <i>Standard Course of Study</i>. Students consistently demonstrate an understanding of conventional written expression by editing various sentence types for correctness, clarity, and style. By inferring, generalizing, drawing conclusions, and making connections between texts, students comprehend with breadth and depth a variety of ninth-grade-level texts. Students can analyze the impact of details and more complex literary elements and devices (such as style, diction, and tone) on the work as a whole. Students can analyze and evaluate purpose, audience, context, and elements of communication particular to expressive, informational, critical, argumentative, and other literary texts.</p>

The standard-setting panelists first drafted these specific achievement level descriptors and they were then reviewed by accountability and curriculum staff. The cut scores and specific achievement level descriptors were then presented to the SBE for approval. The SBE adopted the complete achievement level descriptors below in November 2007.

4.5 Achievement Level Cut Scores

The achievement level cut scores for the North Carolina English I end-of-course tests are shown in the table below.

Table 9: North Carolina End-of-Course Test of English I achievement levels and corresponding scale scores

Course	Program	Level I	Level II	Level III	Level IV
English I	End-of-Course	≤ 137	138–145	146–156	≥ 157

4.6 Achievement Level Trends

Table 10: Achievement level trends for the North Carolina End-of-Course Test of English I

English I	1999	2000	2001	2002	2003	2004	2005	2006	2007
Level I	10.0	9.0	2.5	2.5	1.6	4.0	3.7	3.2	7.1
Level II	31.0	28.3	24.5	21.1	19.6	14.3	14.3	13.8	19.7
Level III	36.0	35.9	40.3	39.0	39.1	42.5	43.0	42.0	43.9
Level IV	23.0	26.7	32.6	37.8	39.6	39.1	38.9	41.1	29.2

4.7 Percentile Ranking

The percentile rank for each scale score is the percentage of scores less than or equal to that score. A percentile is a score or a point on the original measurement scale (see Appendix F for samples of frequency distribution tables). The percentile rank provides relative information about a student's score on a test relative to other students in the norming year. The percentile ranks for the scores on the North Carolina English I test are calculated based on the first operational administration of the tests. The use of percentile rank reporting allows a meaningful comparison to be made among English I scores at the total test score level.

Chapter Five: Reports

5.1 Use of Test Score 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 in English I who takes the North Carolina End-of-Course Test of English I is given a Parent/Teacher Report. This provides the individual student's scale score, achievement level, and percentile. The Parent/Teacher 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.

Students achieving Level III or Level IV are considered to be proficient and to have met the exit standard for English I for graduation as required by State Board of Education policy HSP-N-004 (16 NCAC 6D. 0503).

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 of School Accountability. High school student performance began to be reported in 1998 in the ABCs of 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.

Some schools that do not meet their goals and that have low numbers of students performing at grade level receive assistance from the state. Other schools, where goals have been reached or exceeded, receive bonuses for the certified staff and teacher assistants in that school. Local school systems received their first results under No Child Left Behind (NCLB) in July 2003. Under NCLB, each school is evaluated according to whether or not it met Adequate Yearly Progress (AYP). AYP is not only a goal for the school overall, but also for each subgroup of students in the school. Every subgroup must meet its goal for the school to meet AYP.

AYP is only one part of the state's ABCs accountability model. Complete ABCs results are released in September and show how much growth students in every school made as well as the overall percentage

of students who are proficient. 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 is included in annual reports released nationally on selected subjects. The state also releases state and local SAT scores each summer.

Chapter Six: Descriptive Statistics and Reliability

6.1 Descriptive Statistics for the First Operational Administration of the Test

The third edition of the North Carolina End-of-Course Test of English I was administered for the first time in 2006–07, beginning with the fall semester block. Descriptive statistics for the first operational year and operational administration population demographics are provided below.

6.2 Means and Standard Deviations for the First Operational Administration of the Third Edition of the North Carolina EOC Test of English I

Table 11: Average scale score for the 2006–07 administration of the North Carolina End-of-Course Test of English I

Course	N	Mean	Standard Deviation
English I	111,013	150.03	9.43

6.3 Population Demographics for the First Operational Administration

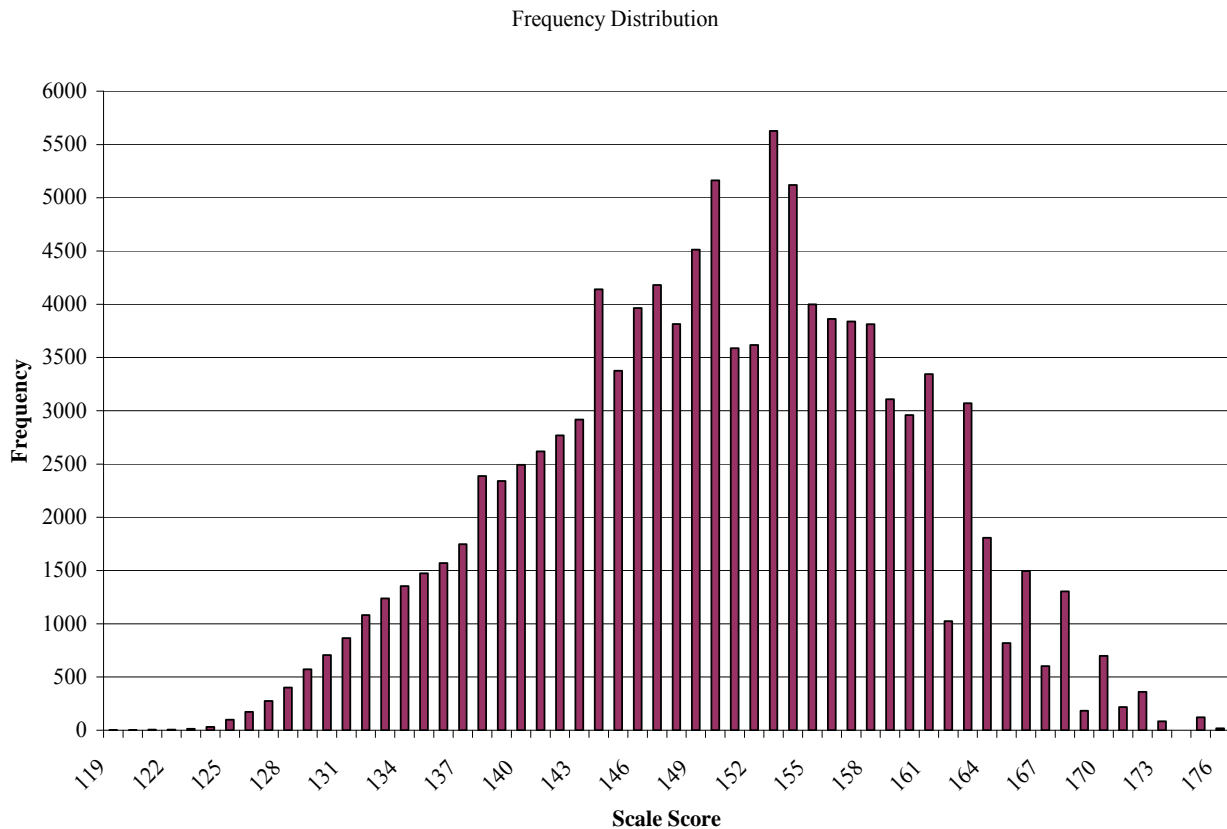
Table 12: Population demographics for the 2006–07 administration of the North Carolina End-of-Course Test of English I

Course	N*	% Female	% Male	% Hispanic	% Black	% White	% Other	% LEP
English I	111,013	49.35	50.65	7.11	30.26	56.88	5.75	3.81

*Percentages are not always based on the total number of participants because not all participants indicated their gender, ethnicity, and/or LEP status.

6.4 Scale Score Frequency Distributions

The following figures present the frequency distributions of the scale scores from the first statewide administration of the North Carolina End-of-Course Test of English I. The frequency distributions are not smooth because of the conversion from raw scores to scale scores. Due to rounding in the conversion process, sometimes two raw scores in the middle of the distribution convert to the same scale score resulting in the appearance of a spike in that particular scale score.

Figure 4: North Carolina End-of-Course Test of English I scale score frequency distribution

6.5 Reliability of the North Carolina End-of-Course Test of English I

Reliability refers to the consistency of a measure when the testing procedure is repeated on a population of individuals or groups. In testing, if use is to be made of some piece of information, then the information should be stable, consistent, and dependable. If any use is to be made of the information from a test, then the test results must be reliable. If decisions about individuals are to be made on the basis of test data, then it is desirable that the test results be reliable and tests exhibit a reliability coefficient of at least 0.85.

There are three broad categories of reliability coefficients 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 is known as an internal consistency coefficient (*Standards for Educational and Psychological Testing*, AERA, APA, NCME, 1985, p.27). An internal consistency coefficient, coefficient alpha, is the metric used to establish reliability for the North Carolina End-of-Course Test of English I.

6.6 Internal Consistency of the North Carolina End-of-Course Test of English I

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

Table 13: Reliability indices averaged across the North Carolina End-of-Course Test of English I forms

Course	Average Coefficient Alpha
English I	0.91

As noted above, the English I test is highly reliable as a whole. In addition, it is important to note that this high degree of reliability extends across gender, ethnicity, LEP status, and disability. Looking at the coefficient alpha for the different groups reveals that averages across all test forms of English I, 8 of 12 values, are at or above 0.90.

Table 14: Reliability indices averaged across the North Carolina End-of-Course Test of English I forms (Gender)

Course	Females	Males
English I	0.91	0.91

Table 15: Reliability indices averaged across North Carolina End-of-Course Test of English I forms (Ethnicity)

Course	Asian	Black	Hispanic	Native American	Multi-Racial	White
English I	0.92	0.88	0.91	0.89	0.90	0.91

Table 16: Reliability indices averaged across the North Carolina End-of-Course Test of English I forms (Other Characteristics)

Course	No Disability	Disability	Not LEP	LEP
English I	0.91	0.88	0.91	0.86

Although the North Carolina Testing Program administers alternate forms of the test, it is not possible to calculate alternate-forms reliabilities on the tests within the context of a natural test setting. Students take the test one time, and only those students who do not achieve Level III are required to retake the test. Thus, the natural population of re-testers has a sharp restriction in range, which would lower the observed correlation. Additionally, North Carolina students are extremely test-wise. Attempting to do a special study on test-retest reliability, where one of the administrations does not have stakes for the student, with this population would give questionable results.

6.7 Standard Error of Measurement

The information provided by the standard error of measurement (SEM) 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 standard error of measurement range for scores on the North Carolina End-of-Course Test of English I is provided in Table 16 below. For students with scores within 2 standard deviations of the mean (95% of the students), standard errors are typically 2 to 3 points. For most of the English I scale scores, the standard error of measurement in the middle range of scores, particularly at the cut point between Level II and Level III, is 2 to 3 points. Scores at the lower and higher ends of the scale (above the 97.5 percentile and below the 2.5 percentile) have standard errors of measurement of approximately 4 to 5 points. This is typical as scores become more extreme due to less measurement precision associated with those extreme scores.

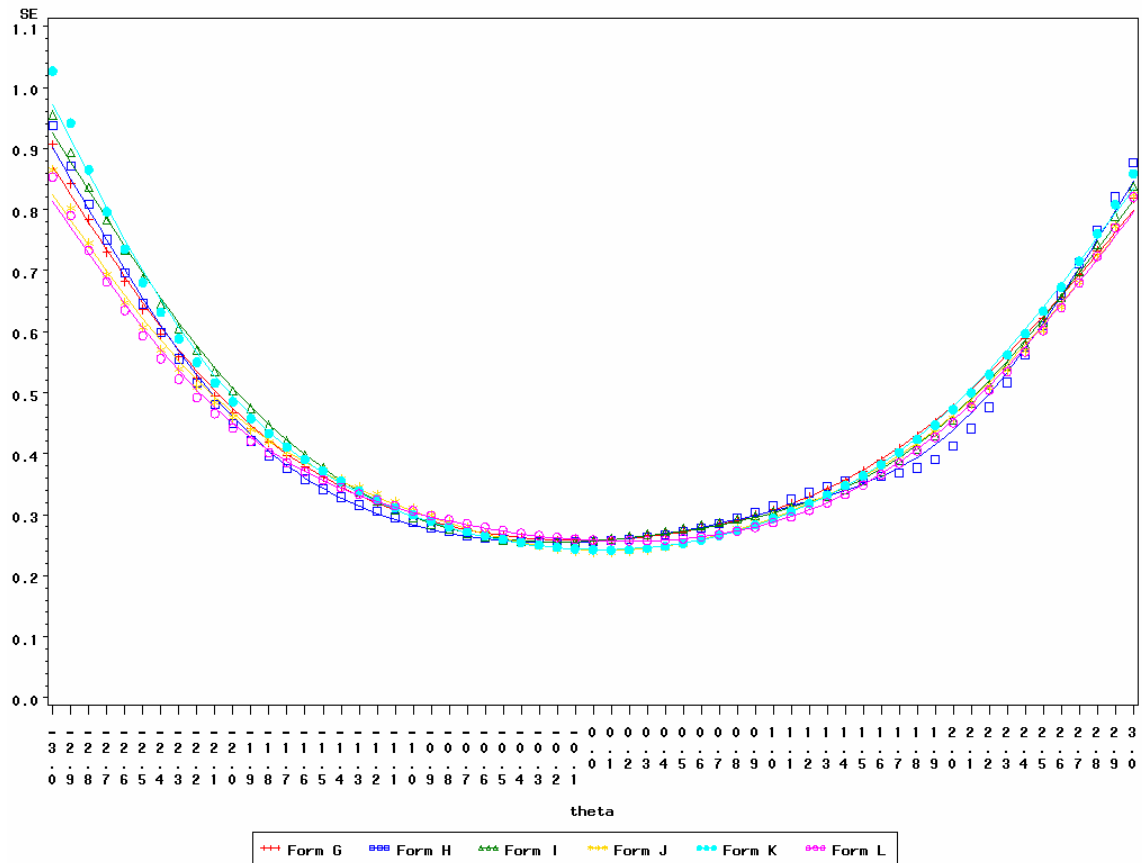
Table 17: Range of standard error of measurement for the North Carolina End-of-Course Test of English I scale scores

Course	Standard Error of Measurement (Range)
English I	2–5

Additionally, standard error curves are presented in the following figures. These are presented on a (0,1) scale on the x -axis representing the θ estimate (the estimate of the test-taker's true ability) for examinees.

Figure 5: Standard errors of measurement on the North Carolina End-of-Course Test of English I forms

English 1 2006–07 Forms GHIJKL: Standard Error Curves (2006–07 Operational Parameters)



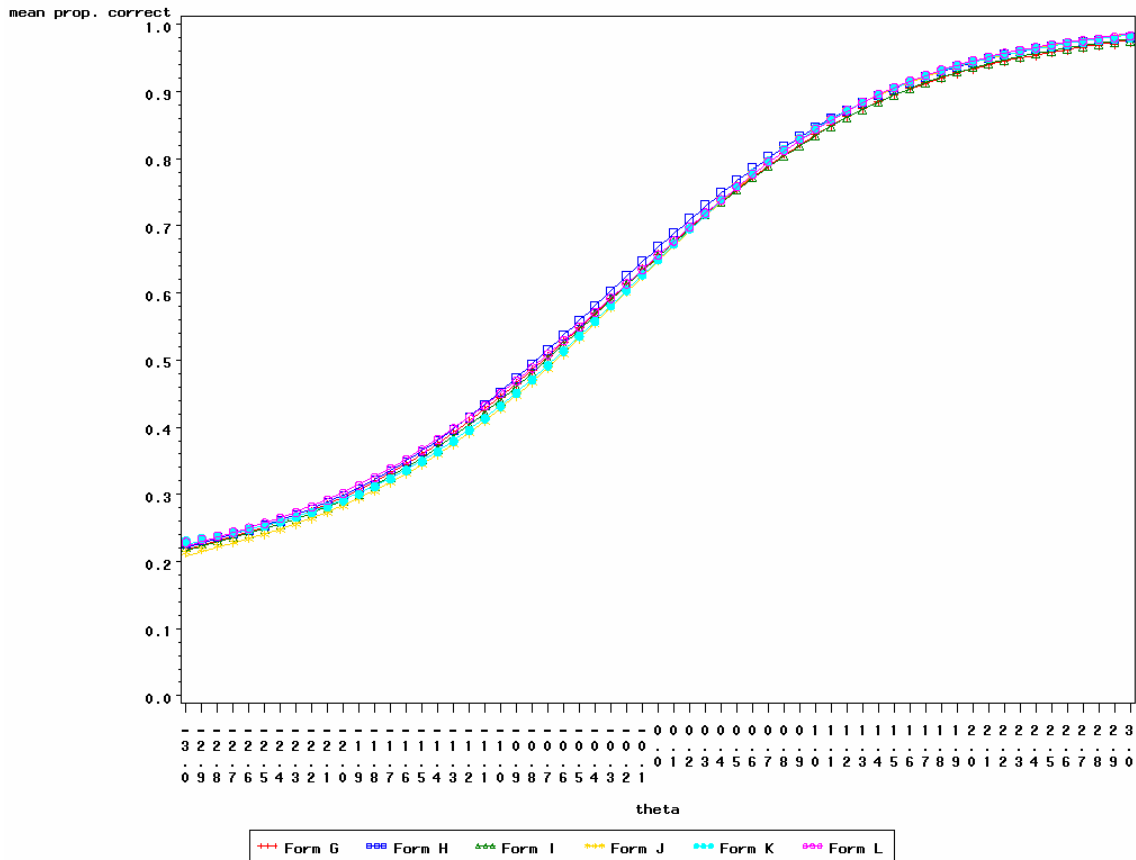
6.8 Equivalency of Test Forms

North Carolina administers multiple forms of each test during each testing cycle. This serves several purposes. First, it allows North Carolina to fully test the breadth and depth of each curriculum. The curricula are extremely rich, and administering a single form that fully addressed each competency would be prohibitively long. Additionally, the use of multiple forms reduces the incidence of one student copying from the test of another student.

The tests are parallel in terms of content coverage by component (i.e., composition versus textual analysis) and by literary environment (i.e., poetry, fiction, informational, expressive nonfiction, argumentative, and critical). The specific questions asked on each form are a random domain sample of the topics in that course’s goals, although care is taken to not overemphasize a particular topic on a single test form.

The tests are statistically equivalent at the total test score level. However, due to the purposively random selection of items tested in each goal, the tests are not statistically equated at the goal level.

Figure 6: Test characteristic curves for the North Carolina End-of-Course Test of English I forms English 1 2006–07 Forms GHIJKL: Test Characteristic Curves (2006–07 Operational Parameters)



For the set of test forms, the test characteristic curves are very nearly coincident for much of the range of θ . Slight variations appear in the test curves at the extremes, as the tests were designed to have maximum sensitivity in the middle of the range of examinee ability.

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 EOC Tests of English I, evidence of validity is provided through content relevance and relationship of test scores to other external variables.

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 North Carolina End-of-Course Test of English I has two sections: composition and textual analysis. Reading selections in the composition section contain errors in mechanics, usage, sentence formation, spelling, organization, purpose, and supporting details. Students are asked to read drafts of student essays and then determine the corrections or changes needed for each essay. If the selection is an analysis of a literary work, it is not necessary for students to have previously read the referenced literary work to answer the items associated with that reading selection. In the textual analysis section of the test, students are asked to read several selections similar to ones they have read in class. After each selection, students are asked to answer questions by analyzing the text.

For test specifications, see Appendix A.

Almost all items are written by North Carolina teachers and other educators. A few of the first round of the third-edition ELA items were written under a contract with a major testing company, but the contract specified that at least half of the items be written by teachers from North Carolina. During the additional field tests, the vast majority of the items were written by North Carolina educators.

Additionally, all items written are reviewed by at least two content-area teachers from North Carolina, and the state's teachers are involved in other aspects of item development and test review. North Carolina educators not only deliver the *Standard Course of Study* every day in their classrooms but are also the most familiar with the way in which students learn and understand the material. Thus, North Carolina teachers are best able to recognize questions that not only match the *Standard Course of Study* for their particular course or grade but also are relevant and comprehensible to the students at that level.

Instructional Validity

DPI routinely administers questionnaires to teachers in an effort to evaluate the validity and appropriateness of the end-of-course tests. Teachers are asked to evaluate the following statements using a five-point scale, with the highest score being “to a superior degree,” and the lowest score being “not at all.”

1. The test content reflects the goals and objectives of the English I curriculum.

2. The test content reflects the goals and objectives of the English I curriculum as it is taught in my school or school system.
3. The items are clearly and concisely written, and the vocabulary is appropriate to the target age level.
4. The content is balanced in relation to ethnicity, race, sex, socioeconomic status, and geographic districts of the state.
5. Each of the items has one and only one answer that is best; however, the distractors appear plausible for someone who has not achieved mastery of the represented objective.

In the most recent administrations, responses to statements reflect that the tests generally met these criteria to a “superior” or “high” degree.

Table 18: Instructional validity of the content of the North Carolina End-of-Course Test of English I

Statement	% indicating to a superior or high degree
1	94.10
2	88.88
3	68.75
4	56.25
5	62.50

7.3 Criterion-Related Validity

Analysis of the relationship of test scores to variables external to the test provides 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 North Carolina EOC Test of English I, 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. The correlation coefficients for the North Carolina EOC Test of English I range from 0.51 to 0.69 indicating a moderate correlation between EOC scale scores and their correlated associated variables.* The table below provides the Pearson correlation coefficients for variables used to establish criterion-related validity for the North Carolina End-of-Course Test of English I.

**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 19: Pearson correlation coefficient table for variables used to establish criterion-related validity for the North Carolina End-of-Course Test of English I

Comparison	Pearson Correlation Coefficient
Teacher Judgment of Achievement Level by Assigned Achievement Level	0.58
Teacher Judgment of Achievement Level by Expected Grade	0.69
Teacher Judgment of Achievement Level by Scale Score	0.61
Assigned Achievement Level by Expected Grade	0.51
Expected Grade by Scale Score	0.54

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 mentioned previously on pages 41–42.
- **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 pages 41–42.
- **Expected Grade:** Teachers were also asked to provide for each student the letter grade that they anticipated the student would receive at the end of the grade or course.
- **English I Scale Score:** The converted raw-score-to-scale-score value obtained by each examinee.

DPI found moderate correlations between scale scores and such variables as teachers’ judgment of student achievement, expected grade, and assigned achievement levels (all measures of concurrent validity). The department also found generally low correlations among these scale scores and variables external to the test, such as gender, limited English proficiency, and disability. The vast majority of the correlations between scale scores and gender or limited English proficient were less extreme than ± 0.10 , and most of the correlations between scale scores and disability status were less extreme than ± 0.30 . None of these relationships approached the levels recorded for the selected measures of concurrent validity. These generalizations held across the full range of forms administered by DPI.

Chapter Eight: Quality Control Procedures

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 laser proof 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) precode 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 check 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 Services. Data are then imported into a file and cleaned. When a portion of the data are in, 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 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 Statewide 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.
Asymptote	An item statistic that describes the proportion of examinees that endorsed a question correctly but did poorly on the overall test. Asymptote for a theoretical four-choice item is 0.25 but can vary somewhat by test..
Biserial correlation	The relationship between an item score (right or wrong) and a total test score.
Common curriculum	Objectives that are unchanged between the old and new curricula
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.
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 identify individual items for further bias review.
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.
Slope	The ability of a test item to distinguish between examinees of high and low ability.

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 the test and the arrangement of objectives.
Threshold		The point on the ability scale where the probability of a correct response is fifty percent. 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

- Camera, W. J. & Echternacht, G. (2000). The SAT I and High School Grades: Utility in Predicting Success in College. Research Notes RN-10, July 2000 (p.6). The College Board Office of Research and Development.
- 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.
- Hambleton, Ronald K. (1983). *Applications of Item Response Theory*. British Columbia: Educational Research Institute of British Columbia.
- 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.
- Muraki, E., Mislevy, R.J., & Bock, R.D. (1991). PC-BiMain: Analysis of item parameter drift, differential item functioning, and variant item performance [Computer software]. Mooresville, IN: Scientific Software, Inc.
- Thissen, D., & Orlando, M. (2001). Item response theory for items scored in two categories. In D. SAS Institute, Inc. (1985). *The FREQ Procedure*. In SAS User's Guide: Statistics, Version 5 Edition. Cary, NC: Author.

Additional Resources

- Anastasi, A. (1982). *Psychological Testing*. New York: Macmillan Publishing Company, Inc.
- Angoff, W.H. (1982). Summary and derivation of equating methods used at ETS (Pp. 55–69). In P.W. Holland & D.B. Rubin, *Test equating*. New York: Academic Press.
- Averett, C.P. (1994). North Carolina End-of-Grade Tests: Setting standards for the achievement levels. Unpublished manuscript.
- 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.
- Bock, R.D., Gibbons, R., & Muraki, E. (1988). Full information factor analysis. *Applied Psychological Measurement*, 12, 261–280.
- Campbell, D.T. & Fiske, D.W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56, 81–105.
- Cattell, R.B. (1956). Validation and intensification of the Sixteen Personality Factor Questionnaire. *Journal of Clinical Psychology*, 12, 105–214.
- Gregory, Robert J. (2000). *Psychological Testing: History, Principles, and Applications*. Needham Heights: Allyn & Bacon.
- 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.
- Hinkle, D.E., Wiersma, W., & Jurs, S. G. (1998). *Applied Statistics for the Behavioral Sciences* (pp. 69–70).
- 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.
- Linn, R.L. (1993). Linking results of distinct assessments. *Applied Measurement in Education*, 6, 83–102.
- Mislevy, R.J. (1992). *Linking educational assessments: Concepts, issues, methods, and prospects*. Princeton, NJ: Educational Testing Service.
- 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.
- North Carolina State Board of Education. (1993). Public School Laws of North Carolina 1994. Raleigh, NC: The Michie Company.
- 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.
- Thissen, D. (1991). *MULTILOG user's guide—Version 6*. Chicago, IL: Scientific Software, Inc.
- Thissen, D. (2001). *IRTLRDIFF v.2.0b: Software for the computation of the statistics involved in item response theory likelihood-ratio tests for differential item functioning*. Unpublished ms.
- Thissen & H. Wainer (Eds), *Test Scoring* (pp. 73–140). Mahwah, NJ: Lawrence Erlbaum Associates.
- Thissen, D., Pommerich, M., Billeaud, K., & Williams, V.S.L. (1995). Item response theory for scores on tests including polytomous items with ordered responses. *Applied Psychological Measurement*, 19, 39–49.
- Traub, R.E. (1994). *Reliability for the social sciences: Theory and applications*. Thousand Oaks, CA: Sage Publications, Inc.
- Williams, V.S.L., Pommerich, M., & Thissen, D. (1998). A comparison of developmental scales based on Thurstone methods and item response theory. *Journal of Educational Measurement*, 35, 93–107.

Appendix A: Testing Code of Ethics

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

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.

.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 B: Test Specifications English I Edition 3

Element	Comments
Purpose of the Test	<p>The North Carolina End-of-Course Test of English I is required by General Statute 115C-174.10 (c) as a component of the North Carolina Annual Testing Program. It is a curriculum-based achievement test specifically aligned to North Carolina’s <i>Standard Course of Study</i> and includes a variety of strategies to measure the achievement of North Carolina students. The purposes of the end-of grade tests are:</p> <ol style="list-style-type: none"> 1. To assure that all high school graduates possess the minimum skills and knowledge necessary to function as a member of society, 2. To provide a means of identifying strengths and weaknesses in the education process in order to improve instructional delivery, and 3. 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	<p>According to State Board of Education policy, the standard for grade-level proficiency shall be a test score at Achievement Level III or above on the English I end-of-course test. Students must score a Level III or higher to meet the exit requirement for high school graduation. Test results are also used to determine school, district, and state adequate yearly progress for the federal requirements per <i>No Child Left Behind</i>.</p>
Curriculum Cycle	<p>Test is based on the 2004 English/Language Arts curriculum (adopted by the State Board of Education). Curriculum cycle revision is every five years. The next curriculum revision is 2009.</p>
Content of the Test	
Subject/Grade	English I/Grade 9
Alignment	<p>The English I test will be comprised of selections that are appropriate reading for students in the ninth grade. Authentic selections are used reflecting what a student at that level might read in a class or on his/her own. Selections are reviewed for appropriateness of language and topic, paying particular attention to potential sources of bias or sensitivity. Each selection must have a clear beginning, middle, and end.</p> <p>The item writing process directs item writers to follow the North Carolina <i>Standard Course of Study</i> (NCSCS) when creating items. Following are the competencies in the NCSCS:</p> <p>Goal 1: The learner will express reflections and reactions to print and non-print text and personal experiences.</p> <p>Goal 2: The learner will explain meaning, describe processes, and answer research questions to inform an audience.</p> <p>Goal 3: The learner will examine argumentation and develop informed opinions.</p> <p>Goal 4: The learner will create and use standards to critique communication.</p> <p>Goal 5: The learner will demonstrate understanding of various literary genres, concepts, elements, and terms.</p> <p>Goal 6: The learner will apply conventions of grammar and language usage.</p>

Dimensionality	The construct of the test is unidimensional, requiring the reporting of a total score for the test.
Weighting	By goals
Obj. not/indirectly measured	The following objectives are not measurable in a multiple-choice format: 2.03: Instruct an audience in how to perform specific operations or procedures. 2.04: Form and refine a question for investigation, using a topic of personal choice, and answer that question.
Miscellaneous remarks	The test consists of two parts: composition and textual analysis. The composition section, part 1, contains four student draft papers with five to eight associated questions for each. The student draft papers represent expressive, informational, argumentative, or critical text modes of writing. Students are asked to read the selections and respond to related items about editing and composition. The textual analysis section, part 2, contains seven literary selections: poetry; informational, fictional, or expressive nonfiction texts; and either an argumentative or critical text. Each selection has six to nine associated items. Each item is aligned to a curriculum objective from the North Carolina English Language Arts <i>Standard Course of Study</i> .
Design	
Population	Students enrolled in English I on the first day of the test administration.
Format	
Item type	Multiple choice: stem with four foils
Special item considerations	Items must be original and unique as well as free of bias (cultural, economic, gender, ethnic, or geographic). The reading level of the items must match the grade level of the test. Distractors must be plausible and the language must be clear and precise.
Delivery Mode	
Mode	Paper and pencil
Accommodations	Braille, large print, one test item per page, Braille writer/slate and stylus (and Braille paper), ccranmer abacus, dictation to a scribe, interpreter transliterator signs/cues test, keyboarding devices, magnification devices, students marks answer in test book, student reads test aloud to self, test administrator reads test aloud, hospital/home testing, multiple testing sessions, scheduled extended time, and testing in a separate room
Number of Items (Total)	80 items

Operational	56
Embedded	This edition of the English I test has 24 embedded items.
By section	The test is formatted in 2 sections: composition and textual analysis
Time Limits	The time limit is 120 minutes with an additional two-minute break and 8 minutes for administrative tasks.
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	
Cognitive Taxonomy	Dimensions of Thinking (Marzano et al.)
Stimulus Materials	The composition section has four selections and the textual analysis section has seven selections. Selections were identified by the contractor's content staff and approved by teacher committees, NCDPI curriculum staff, and NCDPI test development staff.
Other Specifications	None
Cost (Total: development, printing, shipping, etc.)	\$23.54 per student
Psychometric Properties	
P-value	.15≤keep≥.85 .85>reserve>.90 .10<reserve<.15

Biserial Correlation	Keep $\geq .25$.25 > Reserve $\geq .15$
Slope	Keep $\geq .7$.7 > Reserve $> .5$
Asymptote	Keep $\leq .35$.35 < reserve $\leq .45$
Threshold	-2.5 \leq keep \leq 2.5 2.5 < reserve < 3.0 -2.57 > reserve > -3.30
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	Blank piece of paper
Testing Window	Last three weeks of the school year
Scoring	
Methods	Scanned and scored locally (NCDPI provided software)
Raw Scores	0-56
Scale Scores (1 st Administration)	Raw scores are converted to a scale score with a mean of 150 and a standard deviation of 10.
Standard Setting	
Achievement Level Ranges & Descriptors	<p>Achievement Level I: Less than or equal to 137</p> <p>Students performing at this level do not have sufficient mastery of knowledge and skills of the course to be successful at a more advanced level in the content area.</p> <p>Students performing at Achievement Level I demonstrate the need to develop the composition and reading comprehension skills required in the English I North Carolina <i>Standard Course of Study</i>. Students inconsistently identify and correct rudimentary</p>

language convention errors (such as incorrect verb usage, end punctuation errors, double negatives, capitalization errors, and non-standard verb forms). Students show little to no evidence of reading skills and strategies required to comprehend a variety of ninth-grade-level expressive, informational, argumentative, critical, and literary texts.

Achievement Level II: 138-145

Students performing at this level demonstrate inconsistent mastery of knowledge and skills of the course and are minimally prepared to be successful at a more advanced level in the content area.

Students performing at Achievement Level II demonstrate inconsistent application of the composition and reading skills required in the English I North Carolina *Standard Course of Study*. Students inconsistently apply knowledge of grammar and language usage to identify and correct language convention errors in spelling, punctuation, and simple sentence structure. Students show an initial understanding of basic literary devices and elements (such as symbolism, plot structure, and figurative language). Students may also inconsistently apply strategies such as determining the meaning of unfamiliar vocabulary through context clues and identifying the main idea and supporting details. Students show inconsistent literal comprehension of a variety of ninth-grade-level expressive, informational, argumentative, critical, and other literary works.

Achievement Level III: 146-156

Students performing at this level consistently demonstrate mastery of the course subject matter and skills and are well prepared for a more advanced level in the content area.

Students performing at Achievement Level III typically demonstrate composition and reading comprehension skills required by the English I North Carolina *Standard Course of Study*. Students typically demonstrate an understanding of conventional written expression by editing sentences for correctness, clarity, and style. Students can infer, generalize, draw conclusions, and make connections between texts. Students can analyze the impact of details and literary elements and devices (such as characterization, dialect, imagery, symbolism, main idea, purpose, context, and figurative language) on the work as a whole. Students are typically able to comprehend and analyze a variety of ninth-grade-level expressive, informational, argumentative, critical, and other literary texts.

Achievement Level IV: Greater than or equal to 157

Students performing at this level consistently perform in a superior manner clearly beyond that required to be proficient in the course subject matter and skills and are very well prepared for a more advanced level in the content area.

Students performing at Level IV demonstrate a strong command of the composition and reading comprehension skills required by the English I North Carolina *Standard Course of Study*. Students consistently demonstrate an understanding of conventional

	written expression by editing various sentence types for correctness, clarity, and style. By inferring, generalizing, drawing conclusions, and making connections between texts, students comprehend with breadth and depth a variety of ninth-grade-level texts. Students can analyze the impact of details and more complex literary elements and devices (such as style, diction, and tone) on the work as a whole. Students can analyze and evaluate purpose, audience, context, and elements of communication particular to expressive, informational, critical, argumentative, and other literary texts.
Method	Item mapping
Reporting	
Levels of Reporting	Student, school, LEA, state
ABCs	Individual Student Report, Class Roster, School Report Card, District Report Card, State Report Card
NCLB	Adequate Yearly Progress (AYP)
Appropriate Use	Measure of English I curriculum.
History of Development	
Committee Members	Mildred Bazemore, NCDPI Test Development Section Chief Laura Kramer, NCDPI Psychometrician Heather Koons, NCDPI Test Development Content Lead Johnna Faulkner, NCDPI Curriculum English/lalanguage arts lead North Carolina Teachers, committee members representing the regions and diversity of the state
Meeting Minutes	(1) Per Valorie Hargett, Section Chief of ELA 6-12, a committee was not convened to adopt the 2004 ELA curriculum; however, a survey was conducted to determine the clarification needs to the 1999 curriculum. The survey gleaned 693 responses representing 82 local education agencies. Of these, 44 percent provided their names and addresses (attached). (2) The standard setting committee for English I convened on October 15-16 at the NCDPI in Raleigh, North Carolina. Pearson Education Measurement facilitated the standard setting process with the 13-member committee. A final report of the standard setting process was provided by the vendor.

Standard Setting Committee Participants:

Last	First	School	Gender	Ethnicity	Region	Designation
Pate	Mary Frances	HOBBDON HIGH	F	Caucasian	5	General Ed.
Terry	Alice	ORANGE HIGH	F	Caucasian	6	General Ed.
Winters	Kari	PINE FOREST HIGH	F	Caucasian	5	General Ed.
Clifford	Lisa	TUSCOLA HIGH	F	Caucasian	1	General Ed.
Tart	Michael	Craven County Schools	M	Caucasian	5	General Ed.
Poulos	Kelley	JAY M ROBINSON HIGH	F	Caucasian	3	General Ed.
Reilly	Erin	MOUNT TABOR HIGH	F	Caucasian	2	General Ed.
McLean	Terry	LINCOLN TON HIGH	F	Other	3	ESL
Basnight	Renee	WASHINGTON HIGH	F	African American	4	General Ed.
Pearce	Alan	SOUTH CENTRAL HIGH	M	Caucasian	4	General Ed.
Christopher	Tracey	PISGAH HIGH	F	Caucasian	1	General Ed.
Byrd	Sarah	NORTH WILKES HIGH	F	Caucasian	2	General Ed.
Stinson	Amy	NORTH GASTON HIGH	F	African American	3	General Ed.

Gender: Male (M); Female (F)

Region: Western (1); Northwest (2); Southwest (3); Northeast (4); Southeast (5); Central (6)

Appendix C: Item Development Guidelines

Content Guidelines

1. Items must be based on the goals and objectives outlined in the North Carolina *Standard Course of Study* in English Language Arts and written for 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, abbreviation, 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 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, D).
22. Order the answer choices in a logical order. Numbers should be listed in ascending or descending order.
23. Each item written 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 where phrases in the stem (clang associations) are repeated in the foils.
28. Avoid including ridiculous options.
29. Avoid grammatical clues to the correct answer.
30. Avoid specific determiners because 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 4-option multiple-choice item, each correct response should be located at each option position about 25% of the time.
32. The items written 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 D: Test Information Sheet**North Carolina End-of-Course Test of English I**

What are the purposes of the NC Testing Program?

The North Carolina end-of-course tests are required by General Statute 115C-174.10 as a component of the North Carolina Annual Testing Program. As stated, 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.” This English I end-of-course test is one component of the North Carolina Testing Program.

At the LEA level, each student’s English I end-of-course assessment results are included in the student’s permanent records and high school transcript. LEAs shall use results from the assessment as at least twenty-five percent (25%) of the student’s final grade in the course. At the state level, student scores from the English I assessment are used in the computation of school growth and performance composites for the ABCs. At the federal level, English I results are used to meet the *No Child Left Behind Act of 2001* (NCLB) high school test requirements in reading.

According to the State Board of Education (SBE) policy HSP-N-004 (16 NCAC 6D. 0503) students entering the ninth grade for the first time in 2006-07 and beyond are required to meet new exit standards. To meet graduation exit standards, students are required to perform at Achievement Level III (with one standard error of measurement) or above on the English I assessment and four other designated end-of-course assessments.

What is measured by the test?

The English I end-of-course test assesses the goals for English I described in the *North Carolina Standard Course of Study*. The test consists of two separate sections: composition and textual analysis. The composition section, part 1, contains four student draft papers with five to eight associated questions for each, for a total of 28 questions. The student draft papers represent expressive, informational, argumentative, or critical text modes of writing. Students are asked to read the student drafts and respond to related questions about editing and composition.

The textual analysis section, part 2, contains seven literary selections and 52 questions. The selections include poetry; informational, fictional, or expressive nonfiction texts; and either an argumentative or critical text. Each selection is followed by six to nine associated questions. The students are asked to answer related questions in which they must analyze the text for general comprehension as well as author’s craft and strategies.

Table 1: North Carolina End-of-Course Test of English I Measured Content

Goal	Description of Goal
1	The learner will express reflections and reactions to print and nonprint text and personal experiences
2	The learner will explain meaning, describe processes, and answer research questions to inform an audience.
3	The learner will examine argumentation and develop informed opinions.

4	The learner will create and use standards to critique communication.
5	The learner will demonstrate understanding of various literary genres, concepts, elements, and terms.
6	The learner will apply conventions of grammar and language usage.

How is the test administered? A multiple-choice test, the North Carolina End-of-Course Test of English I has 56 operational items and 24 embedded field-test items for a total of 80 items. The embedded field-test items are not included in the student score. The test is administered within a fixed block of time, not to exceed four hours, within the last ten days (traditional) or five days (semester) of the course. Multiple equivalent forms are administered in each classroom to provide information for curriculum evaluation and planning. Test scores are aggregated at the North Carolina Department of Public Instruction to fulfill the state and federal reporting requirements for the accountability systems and the state assessment systems.

How was the test developed? The reading selections were chosen by North Carolina English I language arts educators to reflect the variety of reading done by students as described in the curriculum for English I and to ensure accessibility to a majority of students. The selections chosen are ones that would generally be read by students, would be interesting to the students, and have appropriate content for a reading comprehension test at grade 9. The questions for each selection were written and reviewed by trained North Carolina teachers and educators during the 2004-2005 school year and were field-tested in the 2005-2006 school year. Each field-test question was administered to approximately 1,200 students from randomly selected schools across the state. The English I test was implemented statewide for the first time in the fall of 2006.

What kinds of scores do students receive on the test? The scores on the North Carolina End-of-Course Test of English I are reported as scale scores. Each student's scale score is converted to one of the four achievement level categories (shown below) to provide an interpretation of student performance relative to the English I *Standard Course of Study*. In addition, student scores are reported as percentile ranks, which are generated from a statewide distribution of students who took the test during 2006-07, the first operational year.

Table 2: Achievement Levels for the North Carolina End-of-Course Test of English I

Level I	Level II	Level III	Level IV
Less than or equal to 137	138-145	146-156	Greater than or equal to 157

Level I Achievement Level Descriptor

Students performing at this level do not have sufficient mastery of knowledge and skills of the course to be successful at a more advanced level in the content area.

Students performing at Achievement Level I demonstrate the need to develop the composition and reading comprehension skills required in the English I North Carolina *Standard Course of Study*. Students inconsistently identify and correct rudimentary language convention errors (such as incorrect verb usage, end punctuation errors, double negatives, capitalization errors, and non-standard verb forms). Students show little to no evidence of reading skills and strategies required

to comprehend a variety of ninth-grade-level expressive, informational, argumentative, critical, and literary texts.

Level II Achievement Level Descriptor

Students performing at this level demonstrate inconsistent mastery of knowledge and skills of the course and are minimally prepared to be successful at a more advanced level in the content area.

Students performing at Achievement Level II demonstrate inconsistent application of the composition and reading skills required in the English I North Carolina *Standard Course of Study*. Students inconsistently apply knowledge of grammar and language usage to identify and correct language convention errors in spelling, punctuation, and simple sentence structure. Students show an initial understanding of basic literary devices and elements (such as symbolism, plot structure, and figurative language). Students may also inconsistently apply strategies such as determining the meaning of unfamiliar vocabulary through context clues and identifying the main idea and supporting details. Students show inconsistent literal comprehension of a variety of ninth-grade-level expressive, informational, argumentative, critical, and other literary texts.

Level III Achievement Level Descriptor

Students performing at this level consistently demonstrate mastery of the course subject matter and skills and are well prepared for a more advanced level in the content area.

Students performing at Achievement Level III typically demonstrate composition and reading comprehension skills required by the English I North Carolina *Standard Course of Study*. Students typically demonstrate an understanding of conventional written expression by editing sentences for correctness, clarity, and style. Students can infer, generalize, draw conclusions, and make connections between texts. Students can analyze the impact of details and literary elements and devices (such as characterization, dialect, imagery, symbolism, main idea, purpose, context, and figurative language) on the work as a whole. Students are typically able to comprehend and analyze a variety of ninth-grade-level expressive, informational, argumentative, critical, and other literary texts.

Level IV Achievement Level Descriptor

Students performing at this level consistently perform in a superior manner clearly beyond that required to be proficient in the course subject matter and skills and are very well prepared for a more advanced level in the content area.

Students performing at Level IV demonstrate a strong command of the composition and reading comprehension skills required by the English I North Carolina *Standard Course of Study*. Students consistently demonstrate an understanding of conventional written expression by editing various sentence types for correctness, clarity, and style. By inferring, generalizing, drawing conclusions, and making connections between texts, students comprehend with breadth and depth a variety of ninth-grade-level texts. Students can analyze the impact of details and more complex literary elements and devices (such as style, diction, and tone) on the work as a whole. Students can analyze and evaluate purpose, audience, context, and elements of communication particular to expressive, informational, critical, argumentative, and other literary texts.

Sample Items On the following pages are samples of the types of items that appear on the North Carolina End-of-Course Test of English I. The objective indicates the curriculum objective the item is designed to assess. The thinking skill corresponds to the level of thinking the item requires as defined by the thinking skills taxonomy adapted from *Dimensions of Thinking* by Robert J. Marzano et al. A description of the Thinking Skills used in the North Carolina Testing Program and additional sample items can be found on the NCDPI web site at www.ncpublicschools.org/accountability/testing

In compliance with federal law, including the provisions of Title IX of the Education Amendments of 1972, the Department of Public Instruction does not discriminate on the basis of race, sex, religion, color, national or ethnic origin, age, disability, or military service in its policies, programs, activities, admissions or employment.

Read this selection about a Peace Corps volunteer and answer the questions that follow.

Only in the Peace Corps

by Rachel Peña

There you are sitting on the side of a muddy Paraguayan road, waiting for a bus, when you bite into a broccoli-filled tortilla and realize that you've made a difference.

Tranquilo. All 32 of you will fit on this 15-passenger bus.

Tranquilo. You'll learn Spanish and Guarani in 90 days.

Tranquilo. This isn't even close to as hot as it gets here.

Be *tranquilo*. Be patient. Relax. Take it easy.

I had come to Paraguay with big plans. I had goals. I had an agenda. I had very specifically defined how I would measure my success. I had trouble being *Tranquilo*.

Perhaps it's part of growing up in America, perhaps, it's just me, but patience was a virtue I'd always had trouble understanding. And now, as I prepared for my Peace Corps service in Paraguay, I was becoming more confused than ever. I thought we were here to help. I thought we were here to make a difference. I didn't think we were here to relax.

And yet, every time I turned around: *Tranquilo.* You'll learn Spanish and Guarani in the next 14 days.

Tranquilo. It gets much hotter than this.

Tranquilo. Tranquilo. Tranquilo.

I can't tell you how many times I heard that word during the three months of training. From trainers, from bus drivers, from my host family. In cafes, in the market, in the health center. And always, it was delivered with a calmness and a smile. In a manner that someone of little patience, like myself, could all too easily mistake for smugness.

At the end of the three months I couldn't wait to get to my site. At last, I could get started. I could begin doing all the things that I had told myself I would accomplish.

My assignment was rural health education. The job involved a variety of basic, but important work. I visited schools to teach children about the importance of brushing their teeth and eating healthy foods. I worked with women to educate them about the dangers of cervical cancer. I worked with a Paraguayan nurse to organize and run PAP* campaigns. I also started a community garden with some teachers and their students. I joined other Peace Corps Volunteers and local Paraguayans to build sanitary latrines. I worked with the community to help establish a health clinic.

I was doing stuff. I was accomplishing things, but I couldn't help but feel I could be doing more. If it wasn't for *tranquilo*.

Tranquilo. The supplies will get here.

Tranquilo. The pigs don't know that they are supposed to stay out of the garden.

Tranquilo. It's not that hot.

The work I was most proud of, and the best times I spent in Paraguay, involved the Mothers Club which I helped form soon after I reached my village. Once a week, we would meet at a different mother's house or sometimes at the river to do laundry and bathe while we talked and learned from one another. Those women soon became my family. They warmly welcomed me into their homes, where they treated me as a daughter, a sister, a granddaughter, and a friend.

Naturally, I always had an agenda for our club meetings. A topic for discussion. A list of things I wanted to accomplish. We would talk about ways to improve nutrition

*Papanicolaou: test to detect cervical cancer

and their diets, breast feeding, pregnancy, vaccinations, child survival, first aid, and just about anything else that was related to keeping moms and their kids healthy.

And, of course, my goal was to do more than simply talk with the women. I wanted to see results.

Tranquilo. Things cannot change overnight.

Tranquilo. More people will show up for your next workshop.

Tranquilo. We can talk about that later. Let's have some tea first.

I wanted to see that my community had the lowest rate of cervical cancer in all of Paraguay. I wanted the Ministry of Health to single out my village as the only community in Paraguay with absolutely no intestinal parasites. I wanted every family to change their diets completely to include vegetables like broccoli, which grew like crazy in the community, but none of the mothers knew how to prepare. And I wanted all this now.

Tranquilo. You are trying too hard.

Tranquilo. You will learn Spanish and Guarani yet.

Tranquilo. It gets much hotter.

Every week I would remind my friends in the Mothers Club about, among other things, the many wonderful qualities of broccoli. Its nutritional value and its ease of preparation. Yet, for all my efforts, I felt I was getting nowhere.

One day, about a year into my service, I was waiting for a bus to Asuncion in front of my neighbor's house. It had rained a few days

earlier and the roads were still muddy so it looked as though the bus might never arrive. I paced back and forth and mumbled to myself.

Tranquilo. The bus will be here soon.

I paced back and forth some more.

Tranquilo. I told myself, the bus will be here soon.

Finally, frustrated, I sat down heavily on the ground. A short while later I looked up to see little Maria Brizuela, the daughter of one of the women in my Mothers Club.

Tranquilo. The bus will be here soon, she said.

I managed a slight smile and noticed that she was holding a plate of tortillas that her mother had sent. She sat down next to me and handed me one.

③4 I took one bite and all my delusions of grandeur slipped away. I let go of my timeline and my rigid agenda. Suddenly the heat and the disappointments and the pigs in my garden and the times when nobody showed up for a workshop and the crowded buses and the total frustration with the slower pace were all worth it.

Maria Brizuela's mother had put broccoli in her tortillas.

Maria explained that her mother was cooking broccoli in lots of different foods. She said that her father was even learning to like it.

After a while, Maria returned down the dirt road, and I returned to waiting for the bus.

Tranquilo, I said to myself, it doesn't matter if the bus ever comes.

"Only in the Peace Corps" by Rachel Peña at www.peacecorps.gov.

1. Which of the following **best** describes Rachel’s approach to her work in Paraguay?
- A realistic but uninspired
 - B cautious and thorough
 - C determined but impatient
 - D relaxed and respectful
2. Based on the selection, what does the word *tranquilo* mean?
- A calm
 - B strong
 - C satisfied
 - D realistic
3. Which **best** explains why the author includes the following statement at the beginning of the selection?
- “There you are sitting on the side of a muddy Paraguayan road, waiting for a bus, when you bite into a broccoli-filled tortilla and realize that you’ve made a difference.”
- A to describe the location of her Peace Corps assignment
 - B to emphasize the difficult conditions she faced in Paraguay
 - C to establish the fact that she hoped to improve nutrition in Paraguay
 - D to emphasize the significance of the experience she is about to describe
4. What does the phrase “delusions of grandeur” refer to in paragraph 34?
- A Rachel’s hopes for the future
 - B Rachel’s basic self-confidence
 - C Rachel’s inflated sense of self-importance
 - D Rachel’s feelings of anger

5. How did the experience at the bus stop ***most likely*** affect Rachel's remaining work in Paraguay?
- A She became even more determined to reach all her goals.
 - B She put all her efforts into convincing people to include broccoli in their diet.
 - C She tried to remember that meaningful change occurs gradually.
 - D She continued to think the people's relaxed attitude was her greatest obstacle.
6. Which conflict is ***most closely*** related to the theme of this selection?
- A Rachel vs. the hot climate
 - B Rachel vs. her own attitude
 - C Rachel vs. the Peace Corps
 - D Rachel vs. the Mothers Club

7. What does the repetition of the word *tranquilo* throughout the selection emphasize about Rachel?

- A her knowledge of Spanish
- B her effort to learn Guarani
- C her impatience with Paraguayan culture
- D her lack of effectiveness in Paraguay

End of Set

In compliance with federal law, including the provisions of Title IX of the Education Amendments of 1972, the Department of Public Instruction does not discriminate on the basis of race, sex, religion, color, national or ethnic origin, age, disability, or military service in its policies, programs, activities, admissions or employment.

Read this student rough draft and answer the questions that follow.

To Whom it May Concern:

(1) I was recently in your Home Supply Store in Appleton, and I purchased a Neat 'n' Clean hand vacuum. (2) It was almost \$30.00 including tax. (3) I was sure I would be pleased with this nice looking Neat 'n' Clean when I purchased it. (4) I bought this item as a gift for my mother. (5) She needed something like this item to clean out the inside of her car. (6) When I brought the item home, I knew I had the perfect gift.

(7) I wrapped it, a week later on December 15th, being that I purchased this on December 8th, I gave it to her for her 40th birthday. (8) The box had not been tampered with on the outside when I bought it. (9) She immediately fell in love with this present. (10) I knew that's what she wanted, and I knew I had bought the perfect gift. (11) She even called her sister, bragging about what she got.

(12) The following Saturday morning she opened it up and was very pleased with it's looks. (13) She got out the owner's manual and read step by step how it worked. (14) She plugged it in, turned it on, and it came on. (15) However, there was a slight problem; it would not suction nothing. (16) We, my father and I both took a look at it, and it did the same thing for us. (17) We all together got out the owner's manual and read it together. (18) My dad got out his tools, took it apart and worked on it, but still once it was put back together it did not work properly.

(19) In the owner's manual it had listed all the parts that should be with it, and they are all there. (20) I am returning it, enclosed in the same box it was purchased in, sealed with tape, and the receipt from where I purchased it. (21) Please send me another one that works properly or the exact amount of money I purchased it for.

(22) Again, the product should have worked properly from the start. (23) I am not asking for a whole lot, but I would like to have it replaced. (24) If you do as I asked I will be greatly satisfied with your company. (25) Thanks for your time and effort in this situation.

Thanks again,

Terry Wall

1. What change should be made to correct sentence 12?
- A change *up* to *up*,
 - B change *very* to *real*
 - C change *pleased* to *please*
 - D change *it's* to *its*
2. What change should be made to correct sentence 15?
- A change *However*, to *However*
 - B change *there was* to *there is*
 - C change *problem; it* to *problem, it*
 - D change *nothing* to *anything*
3. What change should be made to correct sentence 16?
- A change *We*, to *We*
 - B change *father* to *father*,
 - C change *I* to *I*,
 - D change *both* to *both*;
4. Which of the following is a run-on?
- A (1) I was recently in your Home Supply Store in Appleton, and I purchased a Neat 'n' Clean hand vacuum.
 - B (7) I wrapped it, a week later on December 15th, being that I purchased this on December 8th, I gave it to her for her 40th birthday.
 - C (11) She even called her sister, bragging about what she got.
 - D (18) My dad got out his tools, took it apart and worked on it, but still once it was put back together it did not work properly.
5. How could sentences 4 and 5 *best* be combined to enhance clarity and coherence?
- A I bought this item as a gift for my mother, who needed something like this to clean the inside of her car.
 - B I bought this item as a gift for my mother, she needed something like this to clean the inside of her car.
 - C A gift for my mother, she needed something like this to clean the inside of her car.
 - D My mother, she needed something like this to clean the inside of her car.

6. How could sentence 17 **best** be written to improve clarity and reduce repetition?
- A We thought for a while, and then we all decided to get out the owner’s manual and read it together.
 - B The owner’s manual—all together, we got it out and read it.
 - C We got out the owner’s manual and read it together.
 - D All together, we got out the owner’s manual and read it together.

7. Which of the following pieces of information is **most helpful** in supporting the writer’s request for a replacement?
- A The writer knew she had the perfect gift for her mother.
 - B The writer gave the hand vacuum to her mother for her fortieth birthday.
 - C The vacuum had no suction.
 - D The writer and her father read the owner’s manual together.

End of Set

In compliance with federal law, including the provisions of Title IX of the Education Amendments of 1972, the Department of Public Instruction does not discriminate on the basis of race, sex, religion, color, national or ethnic origin, age, disability, or military service in its policies, programs, activities, admissions or employment.

*Answers to
English I Textual Analysis and Composition Sample Items*

Passage Title	Question Number	Correct Answer	Objective Thinking Skill	Number
Only in the Peace Corps	1	C	Analyzing	5.01
Only in the Peace Corps	2	A	Analyzing	6.01
Only in the Peace Corps	3	D	Analyzing	4.02
Only in the Peace Corps	4	C	Analyzing	6.01
Only in the Peace Corps	5	C	Generating	5.01
Only in the Peace Corps	6	B	Analyzing	5.01
Only in the Peace Corps	7	C	Analyzing	1.01

Passage Title	Question Number	Correct Answer	Objective Thinking Skill	Number
Neat 'n' Clean Hand Vacuum	1	D	Knowledge	6.02
Neat 'n' Clean Hand Vacuum	2	D	Applying	6.02
Neat 'n' Clean Hand Vacuum	3	C	Applying	6.02
Neat 'n' Clean Hand Vacuum	4	B	Applying	6.01
Neat 'n' Clean Hand Vacuum	5	A	Organizing	6.01
Neat 'n' Clean Hand Vacuum	6	C	Organizing	6.01
Neat 'n' Clean Hand Vacuum	7	C	Evaluating	3.01

Appendix E: Raw-to-Scale Score Conversion Tables

Form	Raw	Scale	G	51	165
G	0	119	G	52	167
G	1	119	G	53	169
G	2	120	G	54	171
G	3	120	G	55	173
G	4	121	G	56	176
G	5	122	H	0	119
G	6	123	H	1	119
G	7	123	H	2	120
G	8	124	H	3	121
G	9	125	H	4	121
G	10	126	H	5	122
G	11	127	H	6	123
G	12	128	H	7	123
G	13	129	H	8	124
G	14	130	H	9	125
G	15	131	H	10	126
G	16	132	H	11	127
G	17	133	H	12	128
G	18	134	H	13	129
G	19	135	H	14	130
G	20	136	H	15	131
G	21	137	H	16	132
G	22	138	H	17	133
G	23	139	H	18	134
G	24	140	H	19	135
G	25	141	H	20	136
G	26	141	H	21	137
G	27	142	H	22	138
G	28	143	H	23	138
G	29	144	H	24	139
G	30	145	H	25	140
G	31	145	H	26	141
G	32	146	H	27	142
G	33	147	H	28	143
G	34	148	H	29	143
G	35	149	H	30	144
G	36	149	H	31	145
G	37	150	H	32	146
G	38	151	H	33	146
G	39	152	H	34	147
G	40	153	H	35	148
G	41	153	H	36	149
G	42	154	H	37	150
G	43	155	H	38	150
G	44	156	H	39	151
G	45	157	H	40	152
G	46	158	H	41	153
G	47	159	H	42	154
G	48	161	H	43	155
G	49	162	H	44	156
G	50	163	H	45	157

H	46	158	I	42	154
H	47	159	I	43	155
H	48	160	I	44	156
H	49	161	I	45	157
H	50	163	I	46	158
H	51	164	I	47	160
H	52	166	I	48	161
H	53	168	I	49	162
H	54	170	I	50	163
H	55	172	I	51	165
H	56	175	I	52	167
I	0	119	I	53	168
I	1	119	I	54	171
I	2	120	I	55	173
I	3	121	I	56	176
I	4	121	J	0	119
I	5	122	J	1	120
I	6	123	J	2	120
I	7	124	J	3	121
I	8	125	J	4	122
I	9	125	J	5	122
I	10	126	J	6	123
I	11	127	J	7	124
I	12	128	J	8	125
I	13	129	J	9	126
I	14	130	J	10	127
I	15	131	J	11	128
I	16	132	J	12	129
I	17	133	J	13	130
I	18	134	J	14	131
I	19	135	J	15	132
I	20	136	J	16	133
I	21	137	J	17	134
I	22	138	J	18	135
I	23	139	J	19	136
I	24	140	J	20	137
I	25	141	J	21	138
I	26	142	J	22	139
I	27	142	J	23	140
I	28	143	J	24	140
I	29	144	J	25	141
I	30	145	J	26	142
I	31	146	J	27	143
I	32	146	J	28	144
I	33	147	J	29	144
I	34	148	J	30	145
I	35	149	J	31	146
I	36	149	J	32	147
I	37	150	J	33	147
I	38	151	J	34	148
I	39	152	J	35	149
I	40	153	J	36	150
I	41	154	J	37	150

J	38	151	K	34	148
J	39	152	K	35	149
J	40	153	K	36	150
J	41	154	K	37	150
J	42	154	K	38	151
J	43	155	K	39	152
J	44	156	K	40	153
J	45	157	K	41	153
J	46	158	K	42	154
J	47	159	K	43	155
J	48	160	K	44	156
J	49	161	K	45	157
J	50	163	K	46	158
J	51	164	K	47	159
J	52	166	K	48	160
J	53	168	K	49	161
J	54	170	K	50	163
J	55	172	K	51	164
J	56	175	K	52	166
K	0	120	K	53	168
K	1	121	K	54	170
K	2	121	K	55	172
K	3	122	K	56	175
K	4	122	L	0	118
K	5	123	L	1	119
K	6	124	L	2	120
K	7	124	L	3	120
K	8	125	L	4	121
K	9	126	L	5	121
K	10	127	L	6	122
K	11	128	L	7	123
K	12	129	L	8	124
K	13	130	L	9	125
K	14	131	L	10	126
K	15	132	L	11	127
K	16	133	L	12	128
K	17	134	L	13	129
K	18	135	L	14	130
K	19	136	L	15	131
K	20	137	L	16	132
K	21	138	L	17	133
K	22	138	L	18	134
K	23	139	L	19	135
K	24	140	L	20	136
K	25	141	L	21	137
K	26	142	L	22	138
K	27	143	L	23	139
K	28	144	L	24	139
K	29	144	L	25	140
K	30	145	L	26	141
K	31	146	L	27	142
K	32	147	L	28	143
K	33	147	L	29	144

L	30	144	L	44	156
L	31	145	L	45	157
L	32	146	L	46	158
L	33	147	L	47	159
L	34	148	L	48	160
L	35	148	L	49	161
L	36	149	L	50	163
L	37	150	L	51	164
L	38	151	L	52	166
L	39	152	L	53	168
L	40	153	L	54	170
L	41	153	L	55	172
L	42	154	L	56	175
L	43	155			

Appendix F: End-of-Course Standard-Setting Report

EXECUTIVE SUMMARY

Introduction

Committees of North Carolina educators were convened on October 15 and 16 by the North Carolina Department of Public Instruction (NCDPI) to set standards on four North Carolina End-of-Course (EOC) tests: Algebra I, Algebra II, Geometry, and English I. Four committees consisting of a total of 52 educators participated in the two-day conference. During this event the panelists (1) became familiar with the examination, (2) clarified the definitions of the performance levels for the four EOC tests, and (3) applied an item mapping procedure to set cut points. The outcomes of the conference are described in this summary.

Panelist Information

The conference consisted of the following four committees meeting simultaneously: Algebra I, Algebra II, Geometry, and English I. Of the 52 panelists, 12 were in the Algebra I committee, 14 were in the Algebra II committee, 14 were in the Geometry committee, and 12 were in the English I committee. All 52 educators provided voluntary demographic information, although not all responded to each question.

Complete demographic and exit survey information from the panelists will be summarized in the Standard-setting Technical Report. A summary of a subset of panelist demographic information is provided in **Table 1**.

Table 1: Percentage of panelists who reported belonging in each demographic category

Committee	Gender		Ethnicity			District Size		
	Female	Male	Black	White	Other	Large	Medium	Small
Algebra I	91.67	8.33	8.33	91.6		33.33	58.33	8.33
Algebra II	85.71	14.29	14.29	78.57	7.14	28.57	42.86	28.57
Geometry	85.71	14.29	7.14	78.57	7.14	28.57	28.57	42.86
English I	83.33	16.67	16.67	75	8.23	33.33	41.67	25

Method and Procedure

The standard-setting conference began on Monday, October 15. The morning of Monday, October 15 was devoted to introductions to the staff, to standard-setting, and to the EOC tests. For this stage of the conference, all panelists met together in one large room.

After the large group meeting, panelists met in their subject specific groups where they became familiar with the EOC assessments by taking a test comprised of representative items from the operational EOCs; the tests were in order from least to most difficult. The committees then began the process of revising performance level descriptors for each EOC subject. Panelists discussed the knowledge, skills, and abilities that differentiated students in each achievement level by referencing the North Carolina *Standard Course of Study*. Performance level descriptors (PLDs) were created for each of the four achievement levels by each EOC subject committee.

Following the creation of PLDs, panelists received training in item mapping, the judgmental process selected by NCDPI for the standard-setting meeting. In this procedure, panelists were asked to identify the item in an ordered item book that is the last item that a threshold student at a given level would be able to correctly answer. Panelists were instructed to identify the last item in an ordered item book that a threshold student at a given level would have a response probability of at least 0.67 of answering correctly.

The ordered item books were the test booklets used by the panelists to familiarize themselves with the EOC test content prior to writing PLDs. Items were ordered from least to most difficult according to the item map. The item map was constructed by finding the scale score that resulted in the response probability .67 (RP67, Huynh, 1998) for each item. The RP67 was determined for multiple-choice items by using the correction for guessing formula

$$RP_{.67} = \frac{(2 + c)}{3},$$

where c is the IRT c -parameter. The ordered item book was arranged in the same manner as the item map, with a single item per page.

Each item map contained the following:

1. Page number
2. A unique item identifier
3. Strand or content category
4. Correct option

Table 2 shows the subjects and number of items in each book.

Table 2: The number of items in each ordered item book

Subject	Number of Items
---------	-----------------

Algebra I	80
Algebra II	80
Geometry	80
English I	84

As noted, items in the ordered item book were sorted using the corrected-for-chance response probability of $(c + 2) / 3$, where c represents the pseudo-chance or lower asymptote of the 3-parameter logistic item response theory model. As Huynh (2006) explains, the value $p = (c + 2) / 3$ maximizes the information of the correct response.

The cut score at each achievement level was determined by computing the median item number across panelists at a given grade level, rounding up when the median fell between two pages. Theta estimates were computed for each item using the RP67 value. The RP67 theta associated with the item on the median page was converted to a scale score using the NC theta to scale score transformation algorithm. The resulting scale score was rounded to a whole number, rounding up for values .4 and higher.

At the beginning of Rounds 2 and 3 panelists were provided with results generated from the previous round to inform their decision making; all results provided to panelists were reported in page numbers. In Round 2, panelists were informed of their individual cut scores and how they compared to the cut scores of other panelists in their small group. At the beginning of Round 3, panelists were provided the updated information based on Round 2 results. Panelists were also provided with the percentage of students that would be classified in each performance level based on their Round 2 cut scores, often referred to as impact data. After panelists made their third round of cuts the median of all panelists in each EOC content area were used to describe the content group cuts cores. Finally, panelists were briefed on the results of their Round 3 ratings.

Results

Performance Level Descriptors

Each EOC subject group created distinct performance level descriptors which reference specific knowledge, skills, and abilities from the North Carolina *Standard Course of Study* associated with each achievement level. The descriptors created during the standard-setting meeting are included in Appendix E.

Cut Scores

Table 3 summarizes the cut scores after the Round 3 final rating. **Table 3** shows the scale score and the page number for each performance level at each grade. The page number is the last item in an ordered item book that a threshold student at a given level would have a response probability of at least 0.67 of answering correctly.

Table 3: Results of panelists' judgments following Round 3

Subject	LEVEL II		LEVEL III		LEVEL IV	
	Scale Score	Median Page	Scale Score	Median Page	Scale Score	Median Page
Algebra I	143	15.5	151	27.5	159	56
Algebra II	148	7	152	21	159	44.5
Geometry	146	10	152	20	159	49
English I	138	6	146	20	155	60

Table 4 shows the percent of students that would be categorized in each performance level based on the median Round 3 final ratings. **Table 5** summarizes those results into the percent that met the standard (passing) and the percent not meeting the standard.

Table 4: The percent of students that would be categorized in each performance level based on the median rating for the Round 3 final ratings

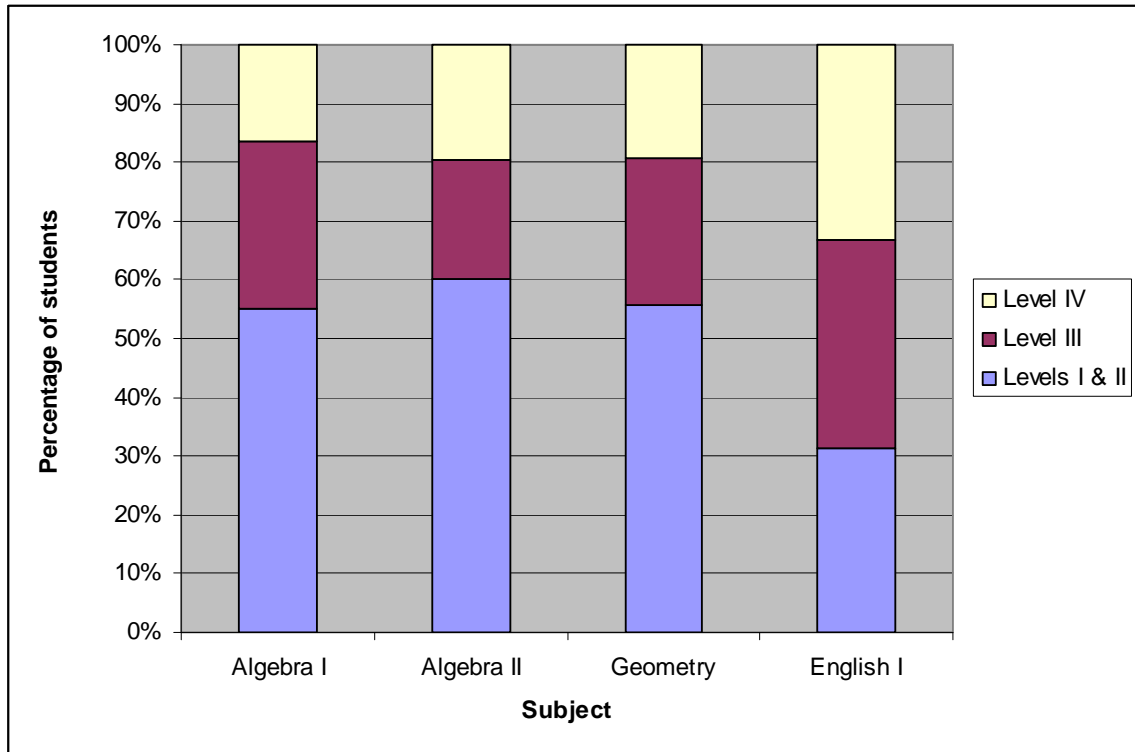
Subject	Levels I & II	Level III	Level IV
Algebra I	55.10	28.30	16.57
Algebra II	59.99	20.41	19.60
Geometry	55.70	25.04	19.26
English I	31.24	35.66	33.10

Table 5: The percent of students that would be categorized in each performance level based on the median rating for the Round 3 final ratings

Subject	Levels I & II (below proficient)	Levels III & IV (proficient and above)
Algebra I	55.1	44.87
Algebra II	59.99	40.01
Geometry	55.7	44.3
English I	31.24	68.76

Figure 1 categorizes student performance by level based on the median Round 3 final ratings.

Figure 1: The percent of students that would be categorized in each performance level based on the median Round 3 final ratings by EOC Subject.



Evaluations

Exit surveys were administered following the completion of standard-setting for each grade. An exit survey was completed by each panelist. The exit surveys consisted of eight questions. These questions and the results for each grade are shown in **Table 6**.

Table 6: Mean response to questions on the exit survey¹

Question	Algebra I	Algebra II	Geometry	English I
1. The method for setting standards, item mapping, was conceptually clear.	4.73	4.91	4.21	4.42
2. I had a good understanding of what the test was intended to measure.	4.77	4.86	4.74	4.67
3. I could clearly distinguish between student performance levels.	4.41	4.59	4.05	4.42
4. After the <u>first</u> round of ratings, I felt comfortable with the standard-setting procedure.	4.27	4.77	3.53	4.12
5. I found the feedback on the ratings of judges compared to other judges useful in setting standards.	4.68	4.95	4.53	4.75
6. I found the feedback on the percent of the students tested that would be classified at each performance level useful in setting standards.	4.73	4.91	4.47	4.42
7. I feel confident that the final cut-score recommendations reflect the performance levels associated with the EOC Test.	4.32	4.77	4.42	4.58

1. Panelists responded using a 1 to 5 scale, with 1 being “**totally disagree**” and 5 being “**totally agree**”.

Note: Responses to items 6 & 7 were made by panelists assuming they had seen accurate impact data

Appendix G: Frequency Distribution for English I Scale Scores

Scale Score	Frequency	Percent	Cumulative Frequency	Cumulative Percent
119	2	0	2	0
120	2	0	4	0
121	7	0.01	11	0.01
122	7	0.01	18	0.02
123	14	0.01	32	0.03
124	33	0.03	65	0.06
125	101	0.09	166	0.15
126	173	0.16	339	0.31
127	275	0.25	614	0.55
128	402	0.36	1016	0.92
129	573	0.52	1589	1.43
130	706	0.64	2295	2.07
131	867	0.78	3162	2.85
132	1082	0.97	4244	3.82
133	1240	1.12	5484	4.94
134	1354	1.22	6838	6.16
135	1474	1.33	8312	7.49
136	1570	1.41	9882	8.9
137	1748	1.57	11630	10.48
138	2389	2.15	14019	12.63
139	2341	2.11	16360	14.74
140	2492	2.24	18852	16.98
141	2620	2.36	21472	19.34
142	2768	2.49	24240	21.84
143	2919	2.63	27159	24.46
144	4140	3.73	31299	28.19
145	3376	3.04	34675	31.24
146	3964	3.57	38639	34.81
147	4182	3.77	42821	38.57
148	3815	3.44	46636	42.01
149	4512	4.06	51148	46.07
150	5164	4.65	56312	50.73
151	3589	3.23	59901	53.96
152	3619	3.26	63520	57.22
153	5628	5.07	69148	62.29
154	5120	4.61	74268	66.9
155	3999	3.6	78267	70.5
156	3862	3.48	82129	73.98
157	3839	3.46	85968	77.44
158	3814	3.44	89782	80.88
159	3110	2.8	92892	83.68
160	2960	2.67	95852	86.34
161	3344	3.01	99196	89.36
162	1024	0.92	100220	90.28
163	3071	2.77	103291	93.04
164	1809	1.63	105100	94.67
165	819	0.74	105919	95.41
166	1495	1.35	107414	96.76

167	603	0.54	108017	97.3
168	1306	1.18	109323	98.48
169	184	0.17	109507	98.64
170	700	0.63	110207	99.27
171	219	0.2	110426	99.47
172	361	0.33	110787	99.8
173	85	0.08	110872	99.87
174	0	0.00	110872	99.87
175	122	0.11	110994	99.98
176	19	0.02	111013	100