The North Carolina Mathematics Tests

Edition 3

Technical Report

Pretest—Grade 3

End-of-Grade Tests (Grades 3–8)

Algebra I End-of-Course Test

Geometry End-of-Course Test

Algebra II End-of-Course Test

CITABLE DRAFT

June 2008

Prepared by:

Mildred Bazemore, Section Chief, Test Development
Laura M.B. Kramer, Ph.D.
Michael P. Gallagher, Ph. D.
North Carolina Department of Public Instruction

Thomas P. Englehart
Robert A. Brown, Ph.D.
Technical Outreach for Public Schools (TOPS)

February 2008

Table of Contents

List of Tables

List of Figures

Chapter One: Introduction

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

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

1.1 Universal Participation

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

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

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

> *The State Board of Education shall develop a plan to create rigorous student academic performance standards for kindergarten through eighth grade and student academic standards for courses in grades 9-12. The performance standards shall align, whenever possible, with the student academic performance standards developed for the National Assessment of Educational Progress (NAEP). The plan also shall include clear and understandable methods of reporting individual student academic performance to parents (G.S. 115C-105.40).*

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

> *H.R. 1 will result in the creation of assessments in each state that measure what children know and learn in reading and math in grades 3-8. Student progress and achievement will be measured according to tests that will be given to every child, every year. …*
> *Statewide reports will include performance data disaggregated according to race, gender, and other criteria to demonstrate not only how well students are achieving overall but also progress in closing the achievement gap between disadvantaged students and other groups of students.*
>
> From: Fact Sheet on the Major Provisions of the Conference Report to H.R. 1, the No Child Left Behind Act

1.2 The North Carolina Testing Program

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

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

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

> *(i) to assure that all high school graduates possess those minimum skills and that knowledge thought necessary to function as a member of society; (ii) to provide a means of identifying strengths and weaknesses in the education*

> *process in order to improve instructional delivery; and (iii) to establish additional means for making the education system at the State, local, and school levels accountable to the public for results.*

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

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

The policy-level generic achievement level descriptors for the Pretest— Grade 3 of Mathematics and the End-of-Grade Mathematics tests administered in grades 3 through 8 are given below:

> **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.

Mathematics End-of-Course tests are administered in Algebra I, Geometry, and Algebra II. The policy-level generic achievement level descriptors for End-of-Course tests are given below:

> **Level I:** 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.
>
> **Level II:** 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.

> **Level III:** Students performing at this level consistently demonstrate mastery of the course's subject matter and are well prepared for a more advanced level in the content area.
>
> **Level IV:** Students performing at this level consistently perform in a superior manner clearly beyond that required to be proficient in the course's subject matter and skills and are very well prepared for a more advanced level in the content area.

The content-specific performance level descriptors are provided for each assessment as Appendix L.

The North Carolina End-of-Grade (EOG) tests include multiple-choice assessments of reading comprehension in grades 3 through 8; mathematics in grades 3 through 8 and 10 (the grade 10 assessment is only for students in Title I schools who have not fulfilled the Algebra I requirement by the 10^{th} grade); and science in grades 5 and 8. There is also a pretest administered at the beginning of the 3^{rd} grade to measure baseline performance in reading comprehension and mathematics. The North Carolina End-of-Course (EOC) tests include multiple-choice assessments of composition and literary analysis in English I and mathematics and mathematical reasoning in Algebra I, Geometry, and Algebra II. In addition to the English and mathematics tests, the North Carolina Testing Program includes science EOC tests in Biology, Chemistry, Physical Science, and Physics; social studies EOC tests in Civics and Economics and U.S. History; writing assessments in grades 4, 7, and 10; the North Carolina Tests of Computer Skills; and alternate and alternative assessments developed to validly measure student abilities in populations who are not able to access the general assessments even with accommodations.

The End-of-Grade tests in grades 3 through 8 mathematics, 3 through 8 reading comprehension, and 5 and 8 science 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 Biology, 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 for end-of-grade tests use a developmental (vertical) scale.

1.3 The North Carolina Mathematics Tests

This technical report for the Third Edition of North Carolina Mathematics Tests discusses tests aligned with the North Carolina Mathematics 2003 *Standard Course of Study* (SCS). Following a five-year revision cycle, the North Carolina State Board of Education adopted the Mathematics *SCS* in March 2003 to replace the 1998 *SCS*. End-of-Grade Tests for grades 3 through 8 were field tested in Spring 2005 and administered operationally for the first time in Spring 2006. The End-of-Grade pretest for grade 3 was field tested at the end of the Summer 2005 and administered operationally the following summer. The End-of-Course Tests in Algebra I, Geometry, and Algebra II were administered as field tests in School Year 2005–2006 and were administered operationally for the first time in School Year 2006–2007.

The purpose of this document is to provide an overview of and technical documentation for the North Carolina Mathematics Tests, 3rd Edition, which include the Pretest— Grade 3, the End-of-Grade Mathematics Tests in grades 3 through 8, and End-of-Course (EOC) Mathematics Tests in Algebra I, Geometry, and Algebra II. Chapter One provides an overview of the North Carolina Mathematics Tests. 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 of 2003, the State Board of Education codified the process used in developing all multiple-choice tests in the North Carolina Testing Program. The development of tests for the North Carolina Testing Program follows a prescribed sequence of events. A 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

[a]Activities done only at implementation of new curriculum
[b]Activities 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 mathematics North Carolina *Standard Course of Study* (NC*SCS*) was in 1998. Following the North Carolina five-year revision cycle, the 2003 revisions "… have been developed through a series of public hearings and the efforts of parents, teachers, educational officials, and representatives of business and industry" (p. 3). State legislation requires alignment with the National Assessment of Educational Progress (NAEP) frameworks. The 2003 revision to the *SCS* was developed to align with the framework for NAEP, which was changed effective for NAEP's 2005 administration. In addition, "The intent of the North Carolina Mathematics *Standard Course of Study* is to provide a set of mathematical competencies for each grade and high school course to ensure rigorous student academic performance standards that are uniform across the state" (p. 2). Hence, alignment with NAEP and rigor were two large themes in the revised *SCS*. The *Standard Course of Study* is available at http://www.ncpublicschools.org/curriculum/mathematics/.

In addition to NAEP, the curriculum review included results from the *Third International Mathematics and Science Study* (TIMMS) and *Principles and Standards of School Mathematics* (National Council of Teachers of Mathematics, 2000).

The North Carolina Mathematics *Standard Course of Study* 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 2003 North Carolina Mathematics *Standard Course of Study*, the goals of mathematics education are for students to develop

(1) strong mathematical problem solving and reasoning abilities;
(2) a firm grounding in essential mathematical concepts and skills, including computation and estimation;
(3) connections within mathematics and with other disciplines;
(4) the ability to use appropriate tools, including technology, to solve mathematical problems;
(5) the ability to communicate an understanding of mathematics effectively; and
(6) positive attitudes and beliefs about mathematics.

The elementary program of mathematics includes knowledge of number facts and computational processes and emphasizes solving problems in a variety of contexts. Middle grades highlight rational numbers and algebra; students will develop fluency in solving multi-step equations and modeling linear functions. High school mathematics includes courses that provide to students the skills and knowledge required for their future. Algebraic and geometric thinking and applied mathematics are essential for all students. Students in North Carolina schools are tested in mathematics in grades 3 through 8. In addition, students taking Algebra I, Algebra II, and Geometry in high school are tested at the end of these courses. Mathematics tests for these grades and courses are designed around the competency goals and objectives found in the NC*SCS*.

2.3 Test Specifications

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

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

(1) percentage of questions from higher or lower thinking skills and classification of each test question into level of difficulty;
(2) percentage of test questions that measure a specific goal and rank order of emphasis for objectives within a goal; and
(3) percentage of questions that require the use of a calculator and percentage that do not allow the use of a calculator.

Test blueprints, specific layouts or "road maps" 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 superordinate goal, and as such, there may be more than one layout. However, at the goal level and in terms of the relative emphasis of the objective coverage, all test blueprints conform to the test specifications.

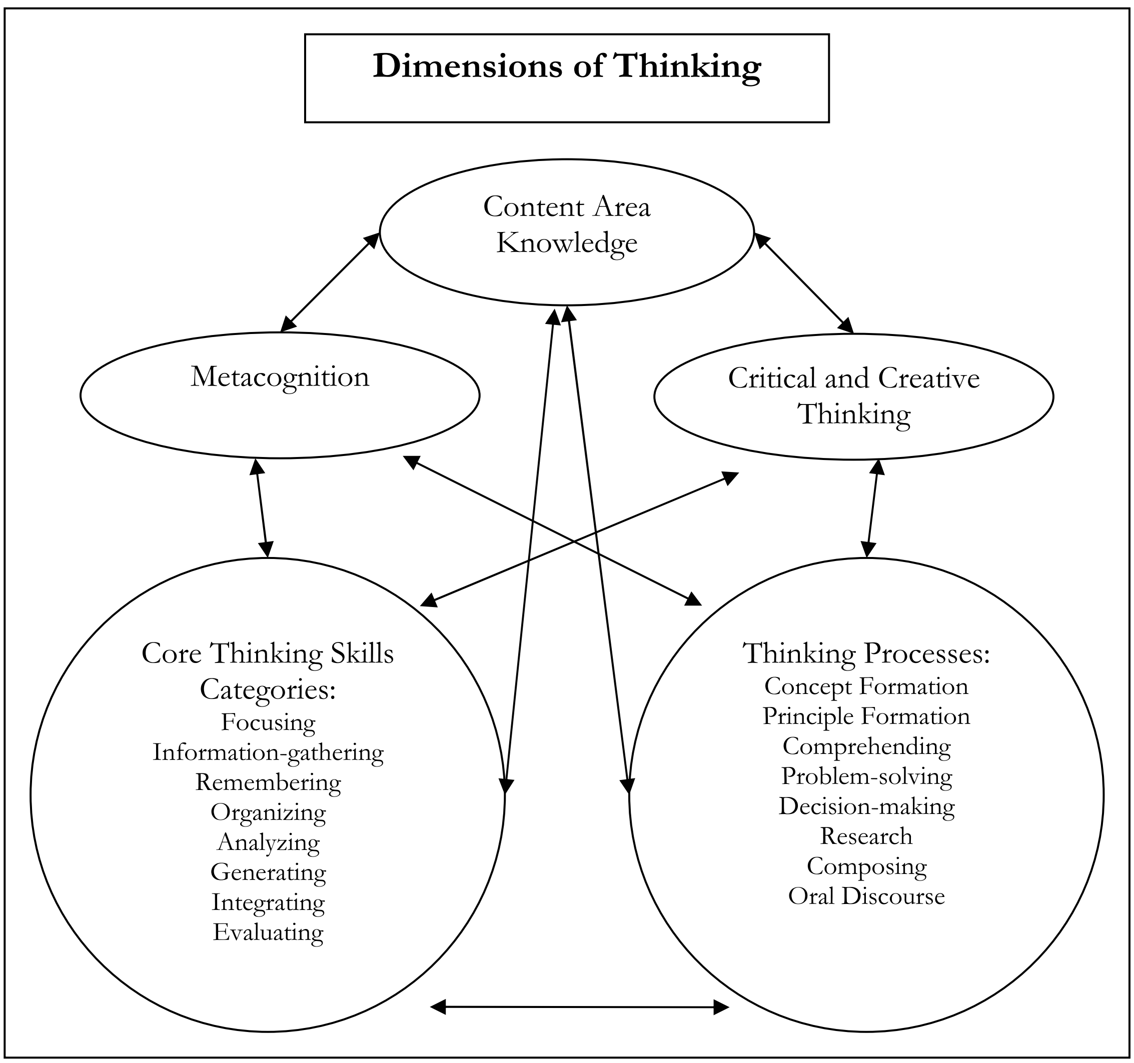
2.4 Item Development

Each item is written to be aligned with a specific objective in the NC*SCS*. Items on the North Carolina EOG and EOC Tests of Mathematics are developed using level of difficulty and thinking skill level. Item writers use these frameworks when developing items. The purpose of the categories is to ensure a balance of difficulty as well as a balance across the different cognitive levels among the items in the North Carolina mathematics tests.

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 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-Grade Tests (adapted from Marzano, et al., 1988)



2.5 Item Format and Use of Manipulatives

Items on the North Carolina mathematics tests are four-foil multiple-choice items. On the end-of-grade mathematics tests, grades 3 through 7, about one-third of the items are calculator inactive items and about two-thirds are calculator active. All items on the grade 8 EOG test and on the EOC tests are calculator active. Apart from what may be presented in an item, students do not have rulers, protractors, or formulas available for any of the math EOG or EOC tests. Graph paper is provided.

2.6 Selection and Training of Item Writers

Once the test blueprints were finalized from the test specifications for the revised edition of the North Carolina mathematics tests, North Carolina educators were recruited and trained to write new items for the state tests. The diversity among the item writers and their knowledge of the current NC*SCS* was 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 NC*SCS* for mathematics. 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 mathematics curriculum, which included information on content and procedural guidelines as well as information on stem and foil development. The item-writing guidelines are included in Appendix A. The items developed during the training were evaluated by content specialists, who then provided feedback to the item writers on the quality of their items.

2.7 Reviewing Items for Field Testing

To ensure that an item was developed to NC*SCS* 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 members, with input from 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 or sensitivity
    - clear statement
    - single problem

- one best answer
- common context in foils
- 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) Diagram/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 2003 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 Mathematics consisted in stand-alone field tests, rather than embedded field tests. (See Table 1.) Because the 2003 *SCS* for grades K–8 was first implemented instructionally in academic year 2004–05, field testing for grades 3 through 8 occurred in Spring 2005, with field testing for the Pretest—Grade 3 deferred until the beginning of academic year 2005–06. The Pretest—Grade 3 content is aligned with the grade 2 content standards and is administered to students during the first three weeks of the grade 3 academic year. Students would not have had an opportunity to learn the material covered in the new grade 2 content standards until the academic year 2004–05; those students became third-graders in 2005–06. The 2003 *SCS* was first implemented instructionally at the high school level in academic year 2005–06. Therefore, field tests for the three mathematics End-of-Course tests were administered in Fall 2005 and Spring 2006.

The first instructional implementation year was transitional in that some curriculum topics, those which had moved from one grade to another, were addressed in both the old and the new grade levels to ensure that students had received instruction in the topic and that the topic did not get "skipped." For example, a topic that was taught in grade 5 in the 1998 *SCS* may have been moved to grade 4 in the 2003 *SCS*. Thus, students entering grade 5 in the transition year would not have yet received instruction in that topic, which likely was a foundation for a topic now introduced in the 5th grade. In the transition year, teachers in the 4th grade would teach the topic as required by the 2003 *SCS*; teachers in the 5th grade would also teach the topic as they did under the 1998 *SCS* so that students could then be prepared to proceed to the topics in the 2003 *SCS*.

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 2003 *SCS*, operational forms aligned to the 1998 *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: Scaling, Equating, and Standard Setting…) is not included in this technical report.

Table 1: Number of items field tested for the North Carolina EOG and EOC Tests of Mathematics (Initial stand-alone field test; 2003 *SCS*-aligned forms only)

Grade or Course	Administration(s)	Number of Forms	Number of Items per Form	Total Number of Items
Pretest—Grade 3	Summer/Fall 2005	8	67	536
Grade 3	Spring 2005	8	82	656
Grade 4	Spring 2005	8	82	656
Grade 5	Spring 2005	8	82	656
Grade 6	Spring 2005	8	82	656
Grade 7	Spring 2005	3	82	246
Grade 8	Spring 2005	3	80	240
Algebra I	Fall 2005/Spring 2006	6	80	480
Geometry	Fall 2005/Spring 2006	6	80	480
Algebra II	Fall 2005/Spring 2006	6	80	480

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. Stratifying variables include

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

The Pretest—Grade 3 field test was administered as a census field test; all eligible students were tested. Table 2 shows the demographic characteristics of the sample for the stand-alone field tests of the Edition 3 mathematics tests.

Beginning with the first operational version of the Edition 3 mathematics tests, field test items are embedded within each form to supplement the item pools. Embedded field test items are grouped into sections. Experimental sections 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) for Pretest—Grade 3 and grades 3–8 end-of-grade tests. Field test population (2005–06) for end-of-course tests

		Gender		Ethnicity						English Language Proficiency Status
Grade or Course	**N**	**% Male**	**% Female**	**% Asian**	**% Black**	**% Hispanic**	**% American Indian**	**% Multiracial**	**% White**	**% LEP** *(Limited English Proficiency)*
Pretest—Grade 3	83,336	51.0	49.0	2.2	27.6	9.4	1.6	3.5	55.8	6.0
Grade 3	15,881	50.7	49.3	2.2	28.3	8.8	1.5	3.3	55.9	5.3
Grade 4	17,983	51.4	48.6	1.9	28.9	8.2	1.4	3.1	56.5	4.6
Grade 5	17,126	50.2	49.8	2.1	29.6	7.9	1.6	3.0	55.9	4.0
Grade 6	16,821	50.7	49.3	2.6	30.1	7.4	2.0	2.8	55.2	3.3
Grade 7	13,313	50.5	49.5	2.0	29.7	7.0	1.6	2.8	57.0	2.8
Grade 8	12,749	50.7	49.3	1.7	29.2	5.1	1.9	2.6	59.6	1.9
Algebra I	10,026	51.4	48.7	2.1	37.0	7.8	0.5	2.9	49.7	4.3
Geometry	6,488	45.2	54.8	2.7	28.8	5.1	1.0	2.9	59.6	1.8
Algebra II	12,717	45.6	54.4	3.0	23.8	4.8	1.3	2.8	64.4	1.2

Notes: The 2005 Pretest—Grade 3 field test administration was a census field test. For grades 3 through 8, Spring 2005 stand-alone field testing was accomplished by drawing demographically representative samples. The N values listed in this table represent examinees taking field test forms aligned with the 2003 North Carolina *Standard Course of Study* (SCS). Additional examinees took test forms containing 1998 *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 for use on an operational North Carolina EOG or EOC Test of Mathematics. 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 contain 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.

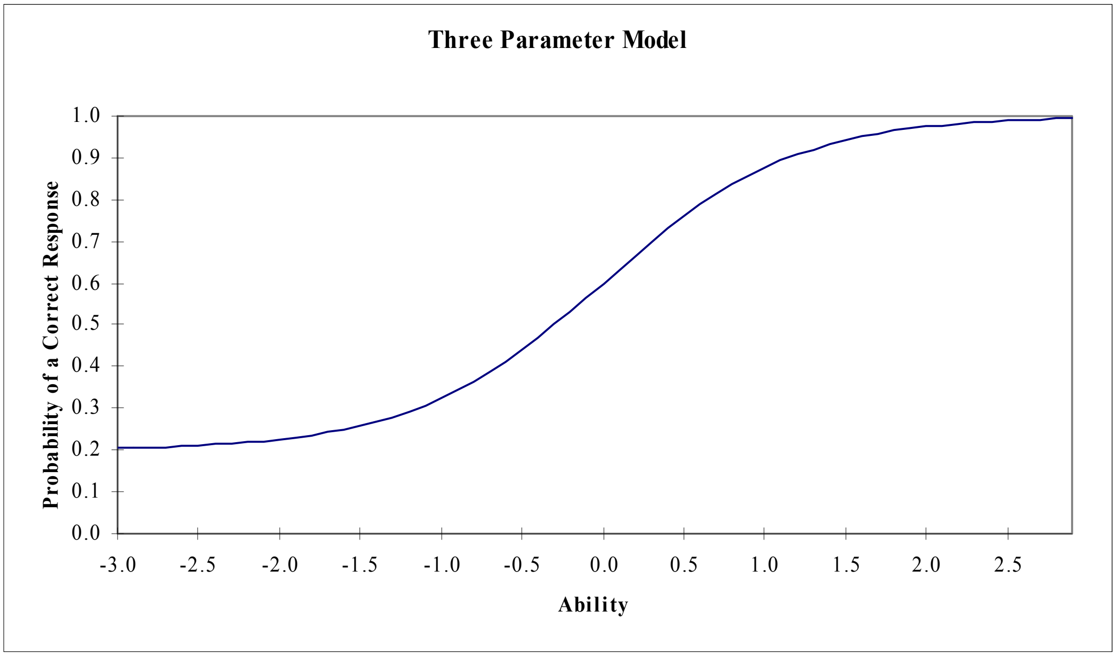
2.11 Classical Measurement Analysis

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 mathematics 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 EOG 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, or "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, or "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*~lognormal(0, 0.5), *b*~N(0, 2), and *c*~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 behaving differently for a particular gender or ethnic group. While the presence or absence of true bias is a qualitative decision, based on the content of the item and the curriculum context within which it appears, DIF can be used to quantitatively identify items that should be subjected to further scrutiny.

In developing the North Carolina mathematics tests, the North Carolina Testing Program staff used the Mantel-Haenszel procedure to examine DIF by examining *j* 2 × 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 NCDPI curriculum specialists.

Items that were flagged by the bias review committee were then reviewed by NCDPI curriculum specialists. If these experts 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 was 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 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 for the End-of-Grade and End-of-Course Tests of Mathematics by grade or course (Pretest and EOG pools from 2005 stand-alone field testing; EOC pools from 2005–06 stand-alone field testing)

Grade or Course	Biserial Correlation	IRT Parameters			P-value			DIF *(Odds-Ratio Logit)*	
		Threshold (b)	*Slope (a)*	*Asymptote (c)*	*Overall*	*Calculator Active*	*Calculator Inactive*	*Ethnicity*	*Gender*
Pretest—Grade 3	0.546	0.220	1.032	0.155	0.5566	0.5652	0.5395	1.092	1.017
Grade 3	0.530	0.289	1.112	0.178	0.5411	0.5362	0.5474	1.108	1.049
Grade 4	0.523	0.474	1.127	0.182	0.4987	0.4971	0.5009	1.091	1.049
Grade 5	0.510	0.347	1.066	0.194	0.5342	0.5337	0.5351	1.068	1.010
Grade 6	0.504	0.648	1.191	0.190	0.4665	0.4700	0.4609	1.079	1.032
Grade 7	0.504	0.619	1.158	0.189	0.4779	0.4840	0.4682	1.041	1.013
Grade 8	0.488	0.720	1.230	0.212	0.4587	0.4587	N/A	1.025	1.001
Algebra I	0.449	0.742	1.044	0.196	0.4636	0.4636	N/A	1.089	0.996
Geometry	0.447	0.854	1.070	0.204	0.4380	0.4380	N/A	1.074	1.016
Algebra II	0.403	1.107	0.989	0.205	0.4036	0.4036	N/A	1.022	1.002

Notes: 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. The grade 8 tests and the three end-of-course tests are entirely calculator active.

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 B. For EOG tests of mathematics, six forms were developed for the first operational administration for the Pretest—Grade 3 and grades 3 through 7. For grade 8, four forms were developed. For the first year of EOC test administration, five forms were developed for Algebra I, four forms for Geometry, and five forms for Algebra II. Additional forms continue to be developed.

2.18 Establishing 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 which may include higher motivation on the operational test, which has stakes for the student; increased or improved background preparation in earlier grades as the curriculum is implemented; 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. Additionally, the end-of-grade math tests have an underlying developmental scale. A conscious decision was made to maintain a monotonically increasing difficulty (i.e., decreasing p-value) across the grade span. The rationale for this was that the material covered in each subsequent grade became more complex. The actual pool p-values generally followed the trend, and the resulting smoothing was relatively minor. 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 and operational p-values

Grade/Subject	p-Value of Item Pool*	p-Value of Forms*	Operational p-Values
Pretest—Grade 3	0.5566	0.5852	0.611
Grade 3	0.5411	0.5392	0.597
Grade 4	0.4987	0.5302	0.596
Grade 5	0.5342	0.5278	0.598
Grade 6	0.4665	0.4999	0.562
Grade 7	0.4779	0.4891	0.570
Grade 8	0.4587	0.4891	0.565
Algebra I	0.4636	0.4498	0.602
Geometry	0.4380	0.4393	0.557
Algebra II	0.4036	0.4055	0.533

* Initial p-values are from stand-alone field testing.

To develop equivalent forms, the test forms were balanced on P+, the sum of the p-values of the items. All calculator active sections within a grade were equated, and all calculator inactive sections within a grade were equated. The sections also have matching or highly similar profiles in terms of numbers of items addressing higher and lower thinking skills and numbers of items categorized as easy, medium, or hard. Finally, 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.

Because of the concerns about student motivation and opportunity to learn on the stand-alone field tests, p-values from the first operational administrations of the tests were also calculated and are included here.

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 a 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 test development staff and consultants. Items that were 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., all area problems in one form were isosceles triangles), 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 Establishing the Test Administration Time

Additional important considerations in the construction of the North Carolina mathematics tests were the number of items to be included and the time necessary to complete the test. Since the tests are power tests, requiring higher-level thinking for many items, students were provided with ample time to complete the test. The *Test Administration Manual* provided test administrators with suggested times, based on the times of 95% of the students finishing the stand-alone field test. See Table 5 below for suggested time on testing (exclusive of distributing materials, reading directions, and so forth).

For grades up through 7, each test has both calculator active and calculator inactive portions. Students take the calculator active and inactive portions in different testing sessions. To reduce test administration irregularities or errors, the calculator active sections are administered before the calculator inactive sections.

Through the 2006–2007 school year, students who were working productively were allowed as much time as they needed to complete the test. Beginning with the 2007-2008 school year, the maximum time allowed for regular students on the End-of-Grade test in grade 8 and the End–of-Course tests in Algebra I, Geometry, and Algebra II was four hours. For the End-of-Grade tests in grades 3–7, the maximum amount of time for the calculator active section is four hours and the maximum amount of time for the calculator inactive section is two and a half hours. This change was enacted after several accounts of test administrations that exceeded a normal school day.

Any student with documented special needs requiring accommodations, such as *Scheduled Extended Time*, of course may exceed these maximum times. Students requiring time beyond the suggested time in the manuals continue to receive 3-minute breaks after every hour of testing.

Table 5: Number of items per test and time allotted by grade and subject

	Number of Items		Suggested Time in Minutes	
Grade / Subject	Calculator Active	Calculator Inactive	Calculator Active	Calculator Inactive
Pretest—Grade 3	42	20	80	40
Grade 3	54	28	135	60
Grade 4	54	28	135	60
Grade 5	54	28	135	60
Grade 6	54	28	135	60
Grade 7	54	28	135	60
Grade 8	80	0	150	N/A
Algebra I	80	0	162	N/A
Geometry	80	0	162	N/A
Algebra II	80	0	162	N/A

Chapter Three: Test Administration

3.1 Test Administration

The North Carolina Pretest—Grade 3 of Mathematics, which measures grade 2 competencies in mathematics, is a multiple-choice test administered to all students in grade 3 within the first three weeks of the school year. The pretest allows schools to establish a baseline to compare individual and group scale scores and achievement levels with the results from the regular end-of-grade test administered in the spring. In addition, a comparison of the results from the pretest and the results from the regular grade 3 end-of-grade test administrations allows schools to measure growth in achievement in mathematics at the third-grade level for the ABCs accountability program. The Pretest—Grade 3 measures the knowledge and skills specified for grade 2 from the mathematics goals and objectives of the 2003 North Carolina *Standard Course of Study*. The pretest is not designed to make student placement or diagnostic decisions.

The End-of-Grade Mathematics Tests are administered to students at grades 3 through 8 as part of the statewide assessment program. The standard for grade-level proficiency is a test score at or above Achievement Level III on both reading comprehension and mathematics tests. Effective with the 2005–2006 school year, the North Carolina End-of-Grade Mathematics Tests are multiple-choice tests that measure the goals and objectives of the mathematics curriculum adopted in 2003 by the North Carolina State Board of Education for each grade. The competency goals and objectives are organized into five strands: (1) number and operations, (2) measurement, (3) geometry, (4) data analysis and probability, and (5) algebra.

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 mathematics end-of-course (Algebra I, Geometry, and Algebra II) tests were developed to provide accurate measurement of individual student knowledge and skills specified in the mathematics component of the North Carolina *Standard Course of Study*. Effective with the 2006–2007 school year, the North Carolina End-of-Course Mathematics Tests are multiple-choice tests that measure the goals and objectives of the mathematics curriculum adopted in 2003 by the North Carolina State Board of Education for each grade. The competency goals and objectives are organized into four strands: (1) number and operations, (2) measurement and geometry, (3) data analysis and probability, and (4) algebra. In schools that follow a traditional calendar, all end-of-course tests are administered within the final 10 days of the course to students enrolled for credit in courses where end-of-course tests are required. For schools that operate under a block or semester schedule, the tests are administered in the last five days of the course.

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 building-level coordinator shall collect, count, and return all test materials to the secure, locked storage area. Any discrepancies are to be reported to the school system test coordinator immediately and a report must be filed with the regional accountability coordinator.

3.5 Student Participation

The Administrative Procedures Act 16 NCAC 6D. 0301 requires that all public school students enrolled in 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 NCAC 6G.0305(g).

Pretest—Grade 3 and End-of-Grade Mathematics Tests (Grades 3 through 8)

All students in membership at grade 3, including students who have been retained at grade 3, are required to participate in the Pretest—Grade 3 Mathematics. All students in membership at grades 3 through 8 are required to participate in the End-of-Grade Mathematics Tests or the corresponding alternate or alternative assessment, as indicated by the student's Individualized Education Program (IEP) or appropriate limited English proficient (LEP) documentation. In the gateway grades (3, 5, and 8), two retest opportunities are available; however, the first test score is used for the purpose of AYP and federal accountability.

Algebra I, Geometry, and Algebra II End-of-Course Tests

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 or LEP documentation, in the final 10 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 in 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 Algebra I, Algebra II, and Geometry) as at least 25 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 2006–2007 and beyond are now required to perform at Achievement Level III (with one standard error of measurement) or above on five required end-of-course (EOC) assessments, one of which is Algebra I, 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

The North Carolina Testing Program currently offers the North Carolina Checklist of Academic Skills (NCCLAS), ***NCEXTEND2***, and ***NCEXTEND1*** as options for meeting the assessment requirements at the state and federal levels. The chart below shows which assessments are available.

Table 6: Available assessments in the North Carolina Testing Program

Grade or Subject	General		Modified Format	Modified Achievement Standards	Alternate Achievement Standards
	Unaccommodated	with Accommodations	NCCLAS	*NCEXTEND2*	*NCEXTEND1*
Pretest—Grade 3	X	X			
Grade 3	X	X	X	X	X
Grade 4	X	X	X	X	X
Grade 5	X	X	X	X	X
Grade 6	X	X	X	X	X
Grade 7	X	X	X	X	X
Grade 8	X	X	X	X	X
Algebra I	X	X	X		X
Geometry	X	X	X		
Algebra II	X	X	X		
OCS*				X	

*The Occupational Course of Study (OCS) is followed by high school students with disabilities for whom the general curriculum is not accessible.

The NCCLAS is an assessment process in which teachers utilize a checklist to evaluate student performance on curriculum benchmarks in the areas of reading, mathematics, and/or writing. 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 (i.e., 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, and/or science. These students have received instruction on the grade-level academic content standards outlined in the NC*SCS* and are held to the same grade-level academic achievement standards.

The ***NCEXTEND2*** tests are based on grade-level content standards for the grade in which the student is enrolled and are challenging for eligible students, but the items may be less difficult than the general assessment and the grade-level academic achievement standards are modified accordingly. These tests are also multiple-choice but only have three foils (response options) rather than four foils as on the general assessments. Eligible students for the ***NCEXTEND2*** tests are identified by the IEP team and meet the criteria outlined below.

- The student's progress in response to high-quality instruction is such that the student is not likely to achieve grade-level proficiency within the school year covered by the IEP.
- The student's disability has precluded the student from achieving grade level proficiency, as demonstrated by objective evidence, (e.g., results from standardized state tests, IQ tests, achievement tests, aptitude tests, and psychological evaluations.)

- Beginning in 2007–2008, the student's IEP must include goals that are based on grade-level content standards and provide for monitoring of the student's progress in achieving those goals.

At the high school level, some of these students will follow the Occupational Course of Study (OCS). The OCS tests are structured in the same way as the end-of-grade ***NCEXTEND2*** tests.

The determination of a significant cognitive disability is one criterion for student participation in the ***NCEXTEND1***. The ***NCEXTEND1*** uses standardized tasks to assess student performance on the NC*SCS* Extended Content Standards. These extended content standards capture the essence of the grade-level content standards but allow for students with disabilities to access the curriculum at a different level. Determination of student proficiency is based on alternate academic achievement standards. The IEP team determines if the disability of a student is a significant cognitive disability; other criteria include the following:

- The student requires extensive and explicit instruction to acquire, maintain, and generalize new reading, mathematics, science, and writing skills for independent living.
- The student exhibits severe and pervasive delays in multiple areas of development and in adaptive behavior (e.g., mobility, communication, daily living skills, and self-care).
- The student is receiving instruction in the grade-level *Standard Course of Study* (*SCS*) Extended Content Standards for the subject(s) in which the students are being assessed. For 2007–2008, this last element was clarified to read "in **ALL** assessed content areas." The revised eligibility requirements clearly state that the ***NCEXTEND1*** is not appropriate for students who receive instruction in any of the general course content standards of the NC*SCS*.

Beginning in 2007–2008, the eligibility requirements were amended to more explicitly define a significant cognitive disability as exhibiting "severe and pervasive delays in **ALL** areas of conceptual, linguistic and academic development and also in adaptive behavior areas, such as communication, daily living skills and self-care."

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 U.S. 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 U.S. 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 with significant medical emergencies and/or conditions 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 medical 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).

3.10 Reporting Student Scores

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

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

3.11 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 EOG and EOC Tests of Mathematics

The North Carolina EOG and EOC Tests of Mathematics 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 counting the number of items he or she answered correctly and then converting the number of correct responses to a developmental scale score. Items are assigned a score of 0 if the student did not answer the item correctly and a score of 1 if the student did answer the item correctly. Software developed at the L.L. Thurstone Psychometric Laboratory at the University of North Carolina at Chapel Hill converts raw scores (total number of items answered correctly) to scale scores using the three IRT 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, grade 3 has six EOG Tests of Mathematics 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 a conditional standard error of measurement associated with it. The raw-to-scale score conversion tables are provided as Appendix J.

The Pretest—Grade 3 and the grades 3 through 8 math tests are on a single developmental (vertical) scale. For the third edition, grade 5 was chosen as the "centering" grade and was scaled to have a mean of 350 and a standard deviation of 10. The procedures for determining the starting mean and standard deviation for the other grades is described below and more fully in Appendix C. Because the EOC Tests of Mathematics are not developmental in nature, the scales were calibrated in the norming year to have a mean of 150 and a standard deviation of 10 for each test; otherwise, the procedures for computing scale scores are the same as for the EOG tests.

4.2 Constructing a Developmental Scale

The basis of a developmental scale is the specification of means and standard deviations for scores on that scale for each grade level. In the case of the North Carolina End-of-Grade Tests of Mathematics, the grade levels ranged from the Pretest—Grade 3 (administered in the fall to students in the third grade) through grade 8. The data from which the scale score means are derived make use of special experimental sections, called linking sections, which were administered to students in adjacent grades. A test section used operationally at the 5th grade would have been embedded into the 6th-grade math test in one of the experimental locations;

the linking items would not count toward the 6th-grade students' scores. It is important to note that no single test version had both its experimental sections populated by off-grade linking material and that the links only extended up, not down, e.g., 6th-grade students may have been administered 5th-grade items, but the 6th-grade students would not have been administered 7th-grade items. The difference in performance between grades on these linking items was used to estimate the difference in proficiency among grades.

The third edition of the North Carolina End-of-Grade Tests of Mathematics used IRT to compute these estimates following procedures described by Williams, Pommerich, and Thissen (1998). Table 7 shows the population means and standard deviations derived from the Spring 2006 item calibration for the North Carolina End-of-Grade Tests of Mathematics. Unlike previous editions of the NC EOG Math Tests, the off-grade linking sections were embedded into operational test forms, rather than spiraled in to the stand-alone field test mix.

Table 7: Population means and standard deviations derived from the Spring 2006 item calibration for the North Carolina End-of-Grade Tests of Mathematics, third edition

Grade	Population	
	Mean	Standard Deviation
Pretest—Grade 3	326.98	12.69
Grade 3	339.44	10.97
Grade 4	345.26	10.24
Grade 5	350.00	10.00
Grade 6	351.45	10.41
Grade 7	353.66	10.15
Grade 8	355.42	9.99

The values for the developmental scale shown in Table 7 are based on IRT estimates of differences between adjacent-grade means and ratios of adjacent-grade standard deviations. BILOG-MG software version 3.0 (Zimowski, Muraki, Mislevy, & Bock, 2002) was used. In BILOG-MG, the lower grade was considered the reference group and thus its population mean and standard deviation were set to 0 and 1, respectively. The values of the mean (μ) and standard deviation (σ) of the higher grade are estimated making use of the item response data and the three-parameter logistic IRT model (Thissen & Orlando, 2001). Table 8 shows the average difference between adjacent-grade means (μ) in units of the standard deviation of the lower grade and ratios between adjacent-grade standard deviations (σ) derived from the Spring 2006 item calibration for the North Carolina End-of-Grade Tests of Mathematics. The values in Table 8 are converted into the final scale, shown in Table 7, by setting the average scale score at grade 5 to be 350.0 with a standard deviation of 10.0 and then computing the values for the other grades such that the differences between the means for adjacent grades, in units of the standard deviation of the lower grade, are the same as those shown in Table 7.

Table 8: Average difference between adjacent-grade means in units of the standard deviation of the lower grade and ratios between adjacent-grade standard deviations derived from the Spring 2006 item calibrations for the North Carolina EOG Tests of Mathematics

Grades	Average Mean (μ) Difference	Average (σ) Ratio	Number of Grade-Pair Forms
Pre 3 – 3	0.982	0.864	6
3 – 4	0.531	0.933	6
4 – 5	0.463	0.977	6
5 – 6*	0.145	1.041	5
6 – 7	0.212	0.975	6
7 – 8	0.174	0.984	4

*For grades 5 and 6, the lowest mean difference (0.05) was dropped from the average. Including the mean difference changes the mean ratio to 0.129 and the standard deviation ratio to 1.05.

Individual runs in BILOG-MG were conducted for each of the grade-pair forms. Grades 3P through 7 had six test forms and grade 8 had four test forms, resulting in a total of 40 adjacent-grade forms. Under the assumption of randomly equivalent groups, the form results were averaged within grade pairs to produce one set of values per adjacent grade. The numbers of replications for each grade pair are also shown in Table 8. Each replication is based on a different short embedded linking section within the operational forms administered in Spring 2006. Depending on the grade and context, linking sections had between 10 and 18 items each. The sample size for each linking section varied from 7,575 to 8,277 students at each grade.

Table 9 shows, for each adjacent-grade pair, the values of the average difference between adjacent-grade means (μ) in units of the standard deviation of the lower grade and ratios of adjacent-grade standard deviations (σ) derived from the Spring 2006 item calibration for the North Carolina EOG Tests of Mathematics for each replication that provided useful data.

Table 9: Replications of the average difference between adjacent-grade means in units of the standard deviation of the lower grade and ratios between adjacent-grade standard deviations derived from the Spring 2006 item calibration for the North Carolina EOG Tests of Mathematics

Grade 3P–3		Grades 3–4		Grades 4–5		Grades 5–6		Grades 6–7		Grades 7–8	
Mean (μ) Difference	(σ) Ratio	Mean (μ) Difference	(σ) Ratio	Mean (μ) Difference	(σ) Ratio	Mean (μ) Difference	(σ) Ratio	Mean (μ) Difference	(σ) Ratio	Mean (μ) Difference	(σ) Ratio
0.958	0.795	0.517	0.922	0.480	0.981	0.155	1.025	0.132	0.924	0.185	0.978
0.943	0.883	0.567	0.899	0.466	0.973	0.156	1.000	0.240	0.917	0.159	0.967
1.124	0.888	0.527	0.944	0.438	0.989	0.050*	1.096	0.253	0.991	0.198	1.015
1.015	0.895	0.432	0.923	0.414	0.990	0.116	1.104	0.206	1.053	0.155	0.975
0.766	0.784	0.503	0.907	0.434	0.910	0.196	1.042	0.195	0.964		
1.083	0.942	0.642	1.002	0.548	1.021	0.101	1.036	0.243	1.001		

*Dropped, as an outlier, from further analyses

4.3 Comparison with the First and Second Edition Scales

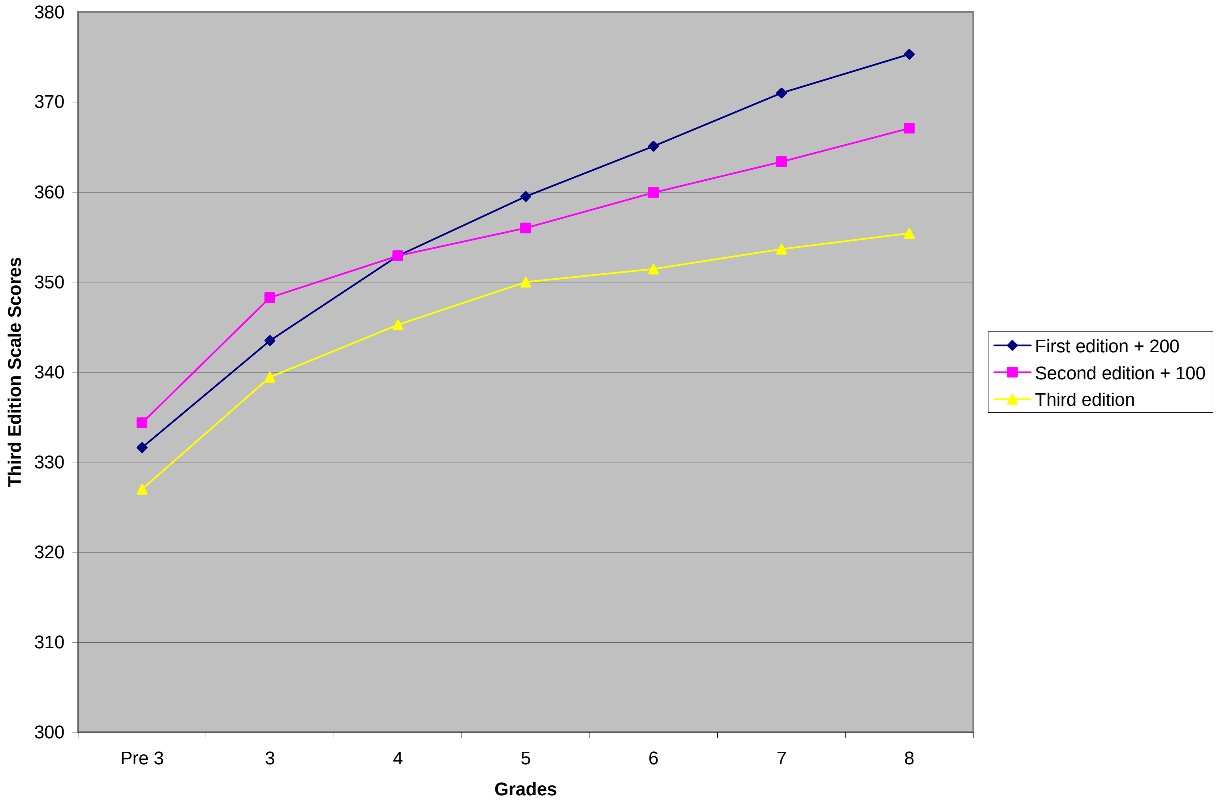
Curriculum revisions in the mathematics *Standard Course of Study*, adopted by the State Board of Education in 2003, resulted in changes in test specifications and a subsequent third edition of the North Carolina EOG and EOC Tests of Mathematics. For a "reasonableness check," the developmental scales from the first and second editions of the North Carolina EOG Tests of Mathematics were compared to the developmental scale created for the third edition of the tests. Table 10 shows a comparison of the population means and standard deviations for the first and second editions with the averages and standard deviations for the scale scores obtained from the operational administration of the third edition.

Table 10: Comparison of the population means and standard deviations for the first, second, and third editions of the North Carolina End-of-Grade Tests of Mathematics

	First Edition (1992)		Second Edition (2000)		Third Edition (2006)	
Grade	**Mean**	**Standard Deviation**	**Mean**	**Standard Deviation**	**Mean**	**Standard Deviation**
Pre 3	131.6	7.8	234.35	9.66	326.98	12.69
3	143.5	11.1	248.27	9.86	339.44	10.97
4	152.9	10.1	252.90	10.65	345.26	10.24
5	159.5	10.1	255.99	12.78	350.00	10.00
6	165.1	11.2	259.95	11.75	351.45	10.41
7	171.0	11.5	263.36	12.46	353.66	10.15
8	175.3	11.9	267.09	12.83	355.42	9.99

To facilitate comparison of the growth between grades among the first, second, and third editions, Figure 4 presents the mean scores plotted together. To place the first, second, and third edition scores on similar scales, a value of 200 was added to the first edition scores and a value of 100 was added to the second edition scores. As shown in Figure 4, the growth of the third edition mean scores is similar to the growth of the second edition mean scores, and quite different from the first edition mean scores. As with the implementation of the 1998 *SCS*, the 2003 curriculum introduced some major topical changes and sweeping changes in emphasis into the content to be taught in the 7th and 8th grades. The tapering-off of growth in the higher grades may be attributed to students entering the higher grades without having had the expected preparation at the lower grades. (This is also the reason the implementation of the edition 3 high school EOC math tests was delayed by one year.)

Figure 4: Comparison of the growth curves for the three editions of the North Carolina EOG Tests of Mathematics



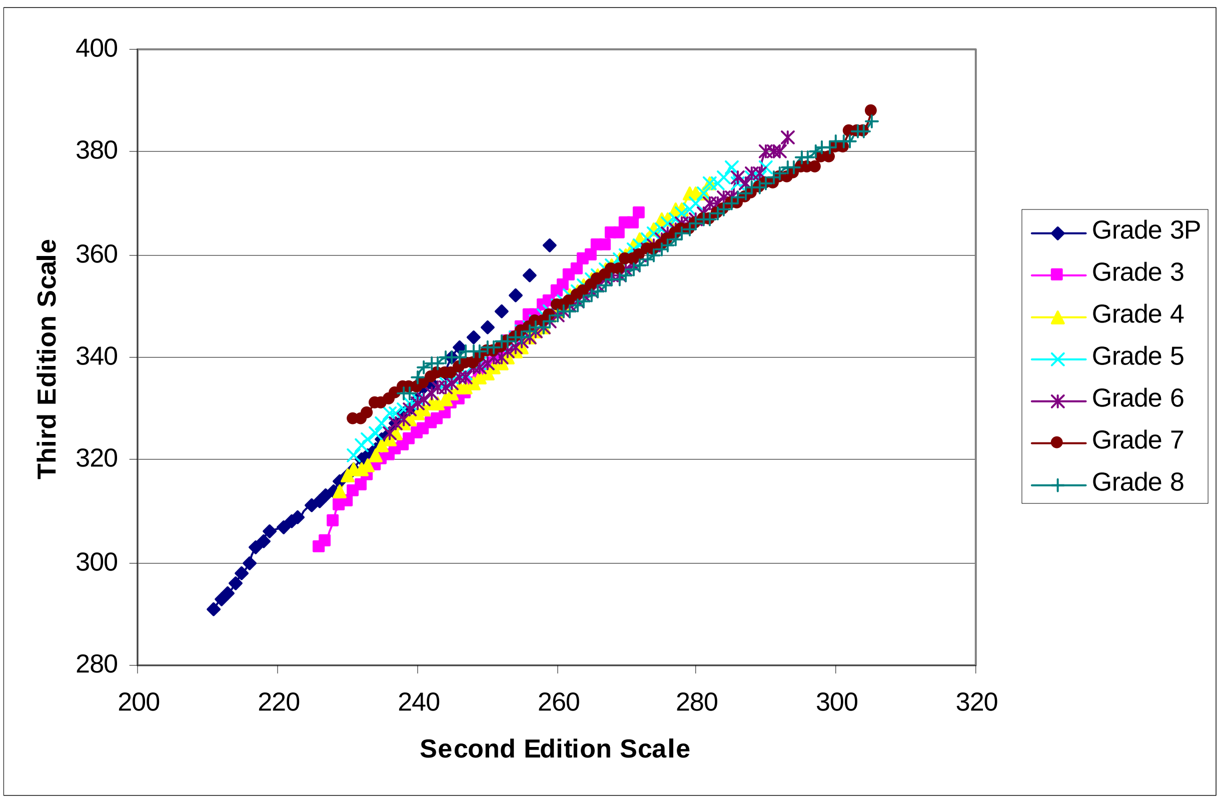
4.4 "Equating" the Scales for the Second and Third Editions of the North Carolina EOG Tests of Mathematics

The third edition scales were statistically moderated to the second edition scales using equipercentile equating methods. While the term "equating" is used throughout this section, the term used is technically inadequate as it should only be applied to tests with parallel content (Mislevy, 1992); strictly speaking, the procedure is "statistical moderating." The third-edition and second-edition tests assess slightly different subdomains of the content area because they reflect revisions to the academic content standards. However, the equipercentile method is an equating process, and therefore it is referred to as "equating" throughout this document and should be understood to be "shorthand" for "the application of equating techniques to a statistical moderation of tests of different content."

The equating process was conducted at each grade, using data gathered in the stand-alone field tests. As noted previously, the year of the stand-alone field tests was considered a transition year between full implementation of the two sets of academic content standards. As such, this was the only time that students would have received instruction in topics covered in both the 1998 and 2003 *SCS*. Thus, even considering the lower motivation frequently cited as a drawback to explicit field testing, this was the best time to be able to collect data on student performance on both sets of content expectations.

Two second-edition forms and either eight (Pretest—Grade 3 and grades 3 through 6) or two (grades 7 and 8) third-edition forms were used in the equating process. The two second-edition forms consisted of the same items but with different ordering. Data from the forms were combined using the following procedure. Within a grade, examinees who took either of the second-edition forms received a raw total score that was then converted to the second-edition scale score. A cumulative frequency distribution was calculated on the scale score data for that grade. Examinees who took a third-edition form received a raw total score, which was then converted to the third-edition scale score for that form and grade. The scale data across forms in a grade were then combined, and a cumulative frequency distribution was computed. Equipercentile equating was then conducted on the cumulative frequency distributions of the second- and third-edition scale scores for each grade, using procedures described in Kolen and Brennan (1995, p. 44). Figure 5 shows the equipercentile equating functions for the Pretest—Grade 3 and grades 3 through 8 obtained using data from the moderating forms.

Figure 5: Equipercentile equating functions between the second and third editions of the North Carolina EOG Tests of Mathematics scales derived from the Spring 2005 explicit field test



One of the research questions to be answered from the application of this equating procedure was to see if it could be reasonable to continue the trend lines for student achievement started in 2002. At that time, each state was required under NCLB to provide an empirical baseline for showing Adequate Yearly Progress (AYP) toward the goal of 100% proficiency in 2014. Any school or district that fails to meet its AYP target each year faces serious penalties. The risk of these sanctions led to this research to see if the scaled scores could be reasonably equated for the purposes of assigning achievement level cut scores, or if completely new standards would need to be set—in effect, starting over from scratch in terms of AYP.

Cut score ranges for the four achievement-level categories developed for the second-edition tests were applied to the third-edition scores using a score translation table derived from the equating study. These ranges appear in Table 11. In a few of the moderating data sets, no student scored as low as the second-edition sample in Level I and no student scored as high as the second-edition sample in Level IV. In addition, there were instances where the upper score of a level was equal to the lower score of the next higher level. Final, smoothed cut scores that addressed these concerns were made in consultation between the NCDPI, Pacific Metrics, and the L.L. Thurstone Psychometric Laboratory.

Table 11. Cut scores for the second and third editions of the North Carolina EOG Tests of Mathematics scale scores for Grades 3P–8 derived from equipercentile equating methods

Grade	Level	Second Edition	Third Edition
Pre 3	I	211-219	291-306
	II	220-229	307-316
	III	230-239	317-331
	IV	240-260	332-358
3	I	218-237	302-322
	II	238-245	323-331
	III	246-254	332-344
	IV	255-276	345-374
4	I	221-239	313-328
	II	240-246	329-334
	III	247-257	335-345
	IV	258-285	346-379
5	I	221-242	319-333
	II	243-249	334-339
	III	250-259	340-349
	IV	260-295	350-388
6	I	228-246	323-336
	II	247-253	337-341
	III	254-264	342-352
	IV	265-296	353-388
7	I	231-249	328-340
	II	250-257	341-346
	III	258-266	347-355
	IV	267-307	356-389
8	I	235-253	330-343
	II	254-260	344-348
	III	261-271	349-357
	IV	272-310	358-389

As was actually expected, the cut scores obtained on the third-edition math tests could not be obtained from the equipercentile equating procedures. In nearly all cases, the cut score that separated Achievement Level I from Achievement Level II represented a raw score at or below the chance level and the cut score that separated Achievement Level II from Achievement

Level III (the proficient – non-proficient cut score) represented a raw score of only about a third of the operational items. Thus, resetting the standards was clearly indicated.

4.5 Setting the Standards

For tests developed under 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, which involves categorizing students into the four achievement levels by expert judges who are knowledgeable of students' achievement in various domains outside of the testing situation and then comparing these judgments to the distributions of students' actual scores. For the North Carolina mathematics tests, 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 new North Carolina tests of mathematics, approximately 100,000 students per grade were placed into categories by approximately 5,000 teachers at each grade. Teachers categorized students who participated in both the field testing and the first year of operational testing into one of the four policy-level generic achievement level descriptors. Fewer than 1% were categorized as not a clear example of any of the achievement levels. Table 12 shows the percentage of students classified into each achievement level by grade or course.

Table 12: Percent of students assigned to each achievement level by teachers

Grade/Subject	Level I	Level II	Level III	Level IV
Pretest—Grade 3 (Fall 2005)	3.92	20.04	53.95	22.09
Grade 3 (Spring 2006)	4.53	18.30	49.63	27.54
Grade 4 (Spring 2006)	4.12	18.76	47.40	29.72
Grade 5 (Spring 2006)	3.26	15.85	48.06	32.82
Grade 6 (Spring 2006)	3.37	15.61	42.99	38.03
Grade 7 (Spring 2006)	3.50	16.30	42.90	37.29
Grade 8 (Spring 2006)	3.54	15.82	42.34	38.29
Algebra I (SY 2006-07)	13.26	24.71	40.11	21.92
Geometry (SY 2006-07)	12.42	25.87	40.01	21.70
Algebra II (SY 2006-07)	11.47	25.43	40.48	22.63

For the third edition of the North Carolina EOG Mathematics Tests, the proportions of students expected to score in each of the four achievement levels were collected in the spring of the first operational administration (Spring 2006). These proportions were applied to the distribution of student scores from the first administration to arrive at one possible set of cut points.

For the third edition of the Pretest—Grade 3 of Mathematics, the contrasting groups data collection was inadvertently omitted from the first administration. The contrasting groups information had fortunately been collected in the (explicit) field test administration in the fall of 2005, and as this data appeared to be consistent with the Spring 2006 operational data collected for the EOG tests, the field test data were used to inform the proportions. For the Pretest—

Grade 3, because this is administered within the first three weeks of the school year, the expert judges for this group of students were the second-grade teachers who had taught the students the previous year. Table 13 shows the number of data points used in this portion of the standard-setting process.

Table 13: Numbers of data points used in the EOG contrasting groups standard setting

Course	N students rated for contrasting groups	N test scores in standard setting frequency distribution	N students in first administration
Pretest—Grade 3	98,176	108,300	109,192
Grade 3	94,962	94,896	104,836
Grade 4	93,047	93,270	102,887
Grade 5	93,916	94,173	103,684
Grade 6	96,808	97,017	106,768
Grade 7	96,544	96,889	106,770
Grade 8	97,535	97,944	107,968

The numbers do not match because not all students received ratings on the 1 to 4 scale; as noted, there were some students who were not a clear example of any of the four categories. Additionally, there were some students who did not have a valid test score and thus were not captured in the frequency distribution. Finally, it is the practice of some districts to complete an answer sheet for all students enrolled in the appropriate grade, even those who are taking an alternate or alternative assessment, as a matter of local bookkeeping; those students would be captured in the overall N for that administration.

When applying the contrasting groups approach to standard setting for the third edition of the North Carolina EOG Mathematics Tests, the scale scores from the first year were displayed in a frequency table from lowest to highest scale score. If the classifications for Grade 3 Math were used as an example, 4.53% of 94,962 would be 4,302 scores. The cumulative frequency distribution allows one to find the scale score below which 4,302 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 and at each grade 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 4,302 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 the students the "benefit of the doubt" at levels I, II, and III, and slightly exceeding the teacher-adjudicated proportion of examinees at level IV. It should be noted that to avoid a further inflation of children categorized as Level IV, the percentage categorized as "not a clear example" was removed from the cut score calculations.

For the third edition of the North Carolina EOC Mathematics Tests, the proportions of students expected to score in each of the four achievement levels were collected in the fall of the first operational year. It is important to note that these proportions were not significantly different when the entire year's data were collected. For Geometry and Algebra II, these proportions then applied to the distribution of student scores from the first operational administration (Fall 2006) to arrive at interim cut scores. For Algebra I, the population of fall examinees was not representative of the entire Algebra I population, so a sample was drawn from the Algebra I fall test-takers to be representative. (Stratifying variables are 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 for all three EOC subjects were used for the entire 2006–2007 school year. Table 14 shows the number of data points used in the interim standard setting for the EOC tests.

Table 14: Numbers of data points used in the EOC interim standard setting

Course	N students rated for contrasting groups	N test scores in interim standard setting frequency distribution	N test scores in the Fall 2006 administration
Algebra I	12,922	13,099	21,829
Geometry	22,670	23,347	23,413
Algebra II	23,109	23,610	23,721

The numbers do not match because not all students received ratings on the 1 to 4 scale; as noted, there were some students who were not a clear example of any of the four categories. As noted, a sample was drawn from the Algebra I fall test-taking population. Additionally, there were some students who did not have a valid test score and thus were not captured in the frequency distribution.

When applying the contrasting groups approach to standard setting for the third edition of the North Carolina EOC Mathematics Tests, scale scores from the fall of the first operational year were displayed in a frequency table from lowest to highest scale score. If the classifications for Geometry were used as an example, 12.42% of 23,413 would be 2,908 scores. The cumulative frequency distribution allows one to find the scale score below which 2,908 students scored. This scale score became the cut-off between Level I and Level II. The process continued for each of the levels in all subjects 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 2,908 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 the students the "benefit of the doubt" 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 categorized as "not a clear example" was removed from the cut score calculations.

Subsequent standard-setting procedures were conducted in the fall of 2006 for the EOG tests and in the fall of 2007 for the EOC tests. 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. The standard-setting reports for both the EOG and EOC mathematics tests are included as Appendices H and I, respectively.

Any standard-setting method is a judgment; however, the goal is to not be capricious. Knowing that any one method will give a different set of cut scores, multiple ways of deriving cut scores were examined in an attempt to arrive at the most appropriate set of cut scores. The item-mapping procedure arrived at a set of cut points that was different from the set obtained by using the contrasting groups method. Additionally, the NCDPI had received a directive from the State Board of Education to increase the rigor of the standards to make the outcome more "NAEP-like."

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 EOG and EOC tests involved looking at the quantitatively derived cut scores as well as considering the policy and practice implications of the cut scores. Because it is so new, this procedure was discussed with the NC Technical Advisory Committee.

4.6 Score Reporting for the North Carolina Tests

Scores from the North Carolina Mathematics Tests 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 or through more authoritative source data collection throughout the school year, such as migrant census, school nutrition data, and so forth. 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 middle of July of each year; disaggregated results are available later in the summer. 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.

4.7 Achievement Level Descriptors

The four policy-level generic achievement descriptors in the North Carolina Testing Program are defined previously, on pages 13–14 of this technical manual. Using these policy-level generic descriptors and the results of the standard-setting, panels of teachers and curriculum experts created a content-based set of achievement level descriptors. After the final standards were approved by the State Board of Education, the achievement level descriptors were reviewed and refined by NCDPI curriculum staff, content experts, and teachers. The goal was to create achievement level descriptors that adequately described what content-specific skills a student should be able to demonstrate to differentiate performance across the four categories without tying student performance to a single test form or administration. The final content-specific achievement level descriptors adopted by the State Board of Education (in 2006 for EOG and 2007 for EOC) are included as Appendix L.

4.8 Achievement Level Cut Scores

The achievement level cut scores for the North Carolina Mathematics Tests are shown in the table below.

Table 15: EOG and EOC Tests of Mathematics achievement levels and corresponding scale scores

Grade/Subject	Level I	Level II	Level III	Level IV
Pretest—Grade 3	≤ 312	313 – 325	326 – 340	≥ 341
Grade 3	≤ 328	329 – 338	339 – 351	≥ 352
Grade 4	≤ 335	336 – 344	345 – 357	≥ 358
Grade 5	≤ 340	341 – 350	351 – 362	≥ 363
Grade 6	≤ 341	342 – 351	352 – 363	≥ 364
Grade 7	≤ 345	346 – 354	355 – 366	≥ 367
Grade 8	≤ 348	349 – 356	357 – 367	≥ 368
Algebra I	≤ 139	140 – 147	148 – 157	≥ 158
Geometry	≤ 138	139 – 147	148 – 157	≥ 158
Algebra II	≤ 138	139 – 146	147 – 157	≥ 158

4.9 Achievement Level Trends

The percentage of students for each of the achievement levels is provided below by grade for selected school years. The years shown include the first year and last year of a test edition and the first year that the test was used in a state-level high-stakes school accountability model.

Table 16: Achievement level trends for Pretest—Grade 3

Grade 3 Pretest	1997	2000	2001	2005	2007
Level I	6.2	3.3	2.0	1.0	6.8
Level II	23.5	19.7	18.9	12.8	28.1
Level III	40.6	41.7	43.4	43.7	47.5
Level IV	29.7	35.3	35.8	43.5	17.6

Table 17: Achievement level trends for Grade 3

Grade 3	1993	1997	2000	2001	2005	2006
Level I	10.7	6.8	5.6	4.2	1.5	7.5
Level II	28.6	23.0	22.6	22.2	12.4	23.8
Level III	39.5	39.6	40.0	43.3	43.4	48.8
Level IV	21.2	30.7	31.8	30.3	42.6	20.0

Table 18: Achievement level trends for Grade 4

Grade 4	1993	1997	2000	2001	2005	2006
Level I	10.0	6.4	2.1	1.2	0.7	8.4
Level II	25.9	19.1	13.4	12.0	6.3	25.7
Level III	44.0	41.9	43.7	46.7	37.9	46.2
Level IV	20.1	32.7	40.8	40.0	55.1	19.7

Table 19: Achievement level trends for Grade 5

Grade 5	1993	1997	2000	2001	2005	2006
Level I	12.1	7.1	3.8	2.2	1.3	8.3
Level II	28.2	19.8	13.3	11.2	7.8	27.8
Level III	28.2	36.2	34.3	36.6	32.3	45.8
Level IV	29.5	36.8	48.6	50.1	58.6	18.2

Table 20: Achievement level trends for Grade 6

Grade 6	1993	1997	2000	2001	2005	2006
Level I	10.5	6.6	4.1	3.3	1.7	9.2
Level II	28.2	20.7	14.9	13.8	8.1	28.2
Level III	41.7	40.5	38.1	40.5	34.2	42.6
Level IV	19.5	32.2	42.9	42.4	56.1	19.9

Table 21: Achievement level trends for Grade 7

Grade 7	1993	1997	2000	2001	2005	2006
Level I	10.5	8.6	4.5	3.2	2.3	11.1
Level II	29.5	20.6	14.8	15.5	12.5	26.6
Level III	38.0	36.9	35.1	33.3	31.2	42.9
Level IV	22.0	34.0	45.6	48.0	54.0	19.5

Table 22: Achievement level trends for Grade 8

Grade 8	1993	1997	2000	2001	2005	2006
Level I	10.4	9.0	4.8	5.3	4.0	13.9
Level II	27.7	22.1	14.6	15.2	11.2	24.9
Level III	41.1	38.4	36.5	36.8	33.9	41.9
Level IV	20.8	30.5	44.1	42.7	50.8	19.4

Table 23: Achievement level trends for Algebra I

Algebra I	1995	1998	2000	2001	2006	2007
Level I	13.9	10.8	9.0	3.2	2.5	12.8
Level II	32.1	27.7	22.1	20.8	17.0	24.7
Level III	40.0	41.9	38.8	44.6	40.0	35.9
Level IV	14.1	19.6	30.1	31.5	40.5	26.6

Table 24: Achievement level trends for Geometry

Geometry	1999	2000	2001	2006	2007
Level I	10.0	9.6	4.7	4.1	9.6
Level II	30.6	30.3	31.4	27.1	26.5
Level III	37.5	36.4	42.1	41.6	37.2
Level IV	20.9	23.6	21.9	27.2	26.7

Table 25: Achievement level trends for Algebra II

Algebra II	1999	2000	2001	2006	2007
Level I	10.0	9.0	2.5	1.6	10.4
Level II	31.0	28.3	24.5	18.1	23.9
Level III	36.0	35.9	40.3	37.7	41.4
Level IV	23.0	26.7	32.6	42.6	24.3

4.10 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. 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 mathematics tests 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 mathematics 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* (see Appendix F) 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 and software 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. School districts who test earlier in the window submit their data to the NCDPI for quality control purposes; those districts are strongly encouraged to not print any reports until the quality control procedures have been completed and the data are certified.

Each student at grades 3–8 who takes the end-of-grade tests is given an Individual Student Report. This single sheet provides information on that student's performance on the reading and mathematics tests. A flier titled, Understanding Your Child's EOG Score, is provided with each Individual Student Report. This publication offers information for understanding student scores as well as suggestions on what parents and teachers can do to help students in the areas of reading and mathematics.

The student report also shows how that student's performance compared to the average scores for the school, the school system, and the state. A four-level achievement scale is used for the tests. A set of global policy-level generic descriptors are used for all subjects:

> Achievement Level I represents insufficient mastery of the subject.
> Achievement Level II is inconsistent mastery of the subject.
> Achievement Level III is consistent mastery and the minimum goal for students.
> Achievement Level IV is superior mastery of the subject.

Additionally, content-specific achievement level descriptors are developed as an outgrowth of the standard-setting process. These are included in Appendix L. Additionally, the appropriate achievement level descriptor is included on the Individual Student Report.

Students achieving at Level III or Level IV are considered to be at or above grade level. Achievement Level III is the level students must score to be considered proficient and to pass to the next grade under the state's Student Accountability Standards for grades 3, 5, and 8.

Beginning in the 2007–2008 school year, the Individual Student Report is being redesigned to provide more feedback to parents and teachers about a student's strengths and weaknesses in the content area. Although the new report only includes information at the goal (superordinate) level, it is actually less prone to misinterpretation or misuse than the older reports, which included information at the objective (subordinate) level. Additionally, the goal summary now is scaled in order to provide somewhat more meaningful interpretations of an individual student's strengths or weaknesses.

Of course it is clearly understood by the NCDPI that reporting at a grain finer than the total test score level is less reliable. Reliability can be bolstered by statistical means such as regression or IRT methods. However, these methods tend to mask individual student profiles by either forcing the student goal-level profile to look more like the state profile, or by flattening the student profile and minimizing any real differences in performance by goal.

In order to attempt to provide meaningful subscore reporting, a student score for each goal is calculated and scaled to have a mean of 10 and a standard deviation of 3, thus, a student goal-level scale score of 10 means the student did about as well on that topic as the rest of the students in the state. A student goal-level scale score below 10 indicates the student did not do as well as other students on that topic, while a student goal-level score above 10 indicates the student did better on that topic than other students in the state. Conditional standard errors are indicated by shaded bands around the point estimate. Strong cautions are included with the report that these scores are less reliable than the total score and that instructional or placement decisions should not be made on the basis of the subscores alone.

5.3 Reporting by Classroom

Classroom rosters can be created from the state-supplied software. These rosters include, for each student, a summary of the information contained on the Individual Student Report. For Algebra I, the classroom roster also provides the numeric "grade" to be factored in to the student's course grade. The default conversion is provided by the state, but the actual conversion used is a local decision. Any district can make its conversion stricter than the state's default, to be more in line with district grading policies.

5.4 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 help from the state. Other schools in which 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 as part of the state's ABCs accountability program. 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 ABCs report to parents.

5.5 Reporting by District

Each district receives its own LEA summary of student performance on the tests that are in the ABCs accountability model as well as information on how the LEA performed in terms of AYP.

5.6 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 Tests

The third editions of the EOG Tests of Mathematics were administered for the first time in the spring of 2006; the new EOC Tests of Mathematics were administered for the first time in the 2006-07 school year (fall and spring block schedule, and traditional calendar); and the Pretest—Grade 3 of Mathematics was administered for the first time in the fall of 2006. Descriptive statistics for the North Carolina Tests of Mathematics' first operational year and operational administration population demographics are provided below.

6.2 Means and Standard Deviations for the First Operational Administration

Table 26: Descriptive statistics by grade or course for the first administration of the North Carolina Tests of Mathematics

Grade	N	Scale Score Mean	Scale Score Standard Deviation	Average p-Value of Tests
Pretest—Grade 3 (Fall 2006)	108,501	329.7	11.4	0.611
Grade 3 (Spring 2006)	104,205	343.2	9.7	0.597
Grade 4 (Spring 2006)	102,306	348.9	9.5	0.596
Grade 5 (Spring 2006)	103,067	353.7	9.2	0.598
Grade 6 (Spring 2006)	106,036	354.9	9.7	0.562
Grade 7 (Spring 2006)	105,764	357.8	9.6	0.570
Grade 8 (Spring 2006)	106,866	359.2	9.2	0.565
Algebra I (SY 2006–07)	116,209	150.9	10.2	0.602
Geometry (SY 2006–07)	80,300	151.1	9.6	0.557
Algebra II (SY 2006–07)	73,332	150.6	9.5	0.533

6.3 Population Demographics for the First Operational Administration

Table 27: Population demographics for the first administration of the North Carolina EOG and EOC Tests of Mathematics

Grade / Subject	% Male	% Female	% American Indian	% Black	% Hispanic	% White	% Other	% LEP	% SWD
Pretest—Grade 3	50.7	49.3	2.0	29.8	10.9	51.8	5.5	5.1	6.3
Grade 3	50.5	49.5	1.5	25.9	9.3	57.6	5.7	6.6	14.4
Grade 4	50.6	49.4	1.6	25.8	8.9	58.4	5.3	5.4	14.3
Grade 5	51.1	48.9	1.5	26.6	8.2	58.6	5.1	5.4	14.2
Grade 6	51.0	49.0	1.6	28.0	7.5	58.2	4.7	3.8	13.2
Grade 7	50.8	49.2	1.5	28.3	6.9	58.9	4.4	3.3	12.7
Grade 8	50.4	49.6	1.6	28.3	6.5	59.6	4.0	3.1	12.8
Algebra I	50.0	50.0	1.2	30.5	7.0	56.7	4.6	3.9	9.1
Geometry	46.6	53.4	1.0	27.0	5.6	61.5	5.0	2.5	4.4
Algebra II	46.6	53.4	1.0	25.5	4.9	63.4	5.1	2.0	3.7

6.4 Scale Score Frequency Distributions

The following figures present the frequency distributions of the developmental scale scores from the first statewide administration of the North Carolina EOG and EOC Tests of Mathematics. 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 6: Math Scale Score Frequency Distribution Pretest—Grade 3

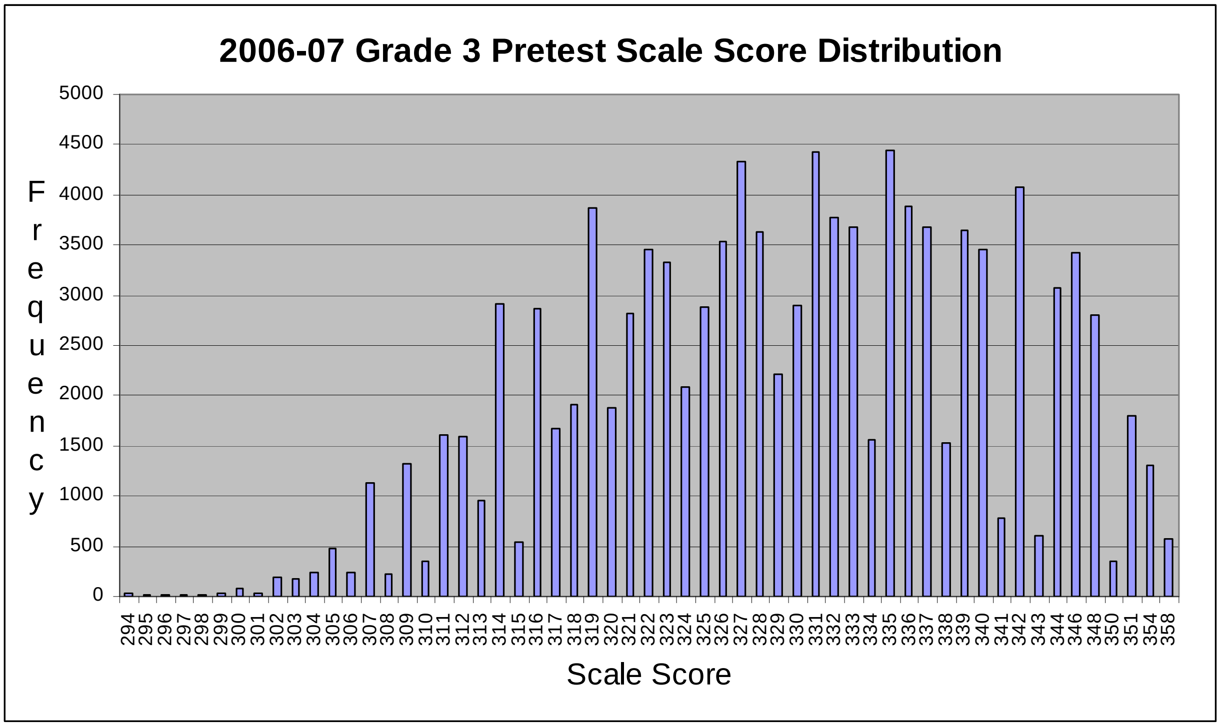


Figure 7: Math Scale Score Frequency Distribution Grade 3

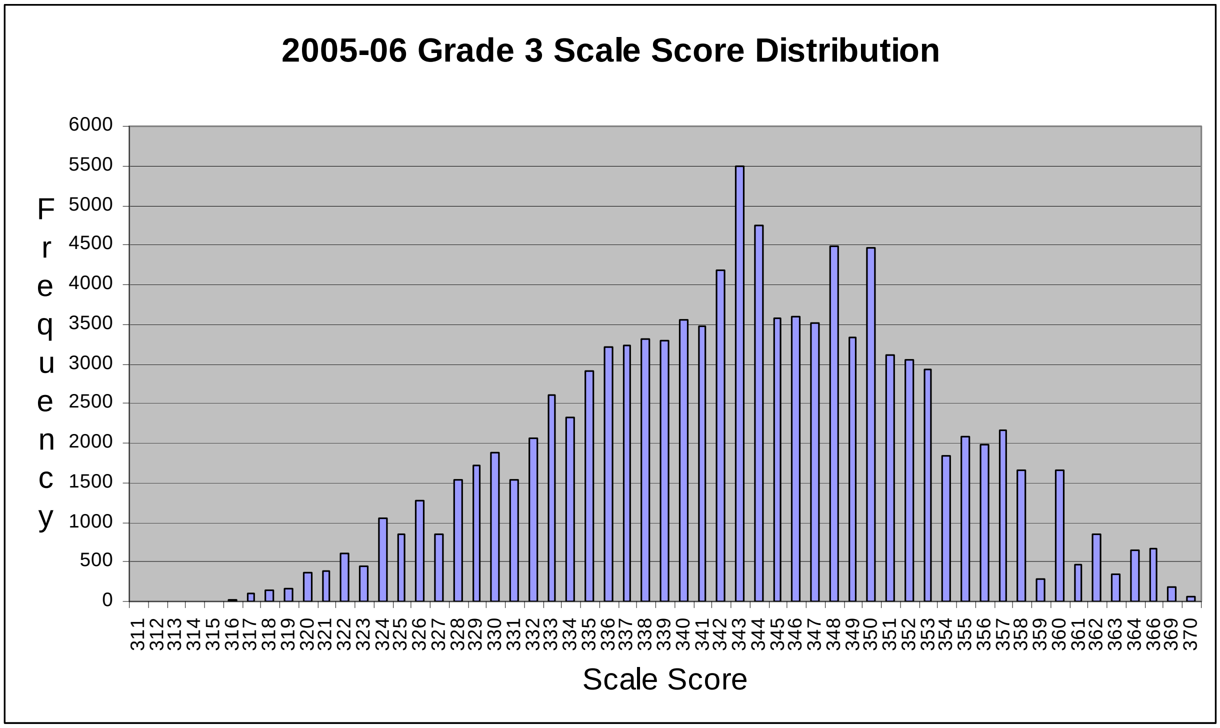


Figure 8: Math Scale Score Frequency Distribution Grade 4

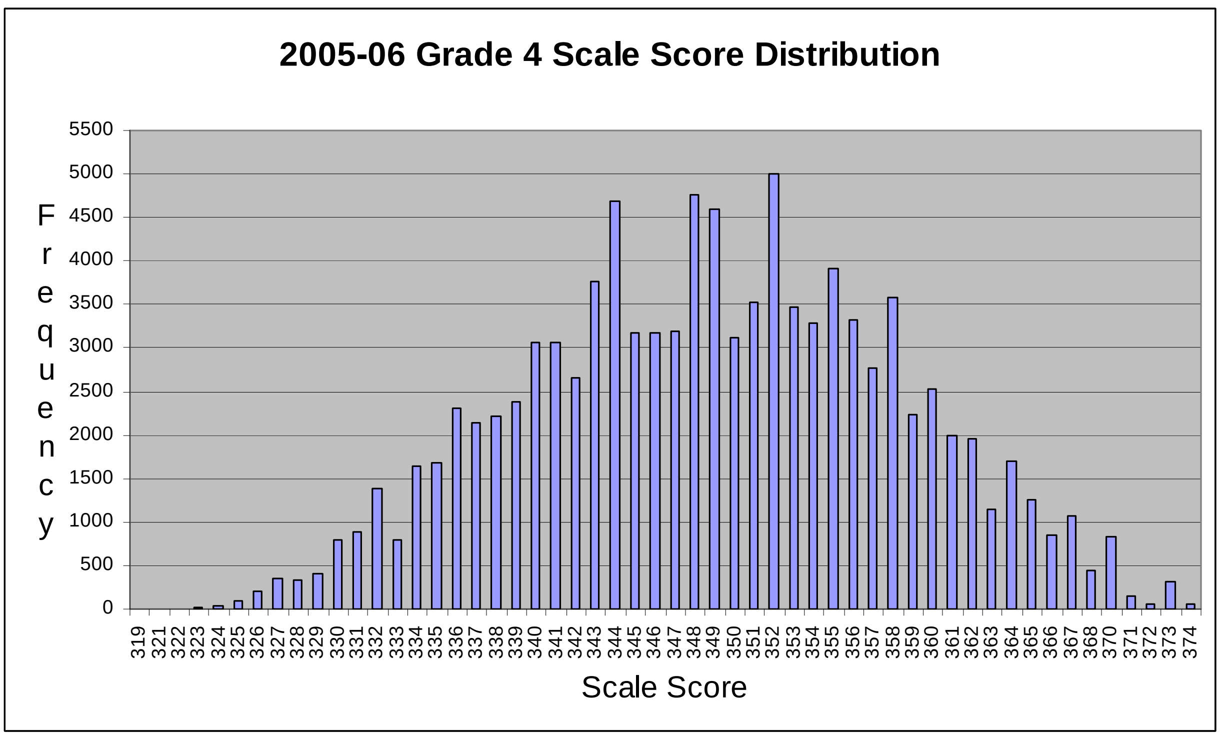


Figure 9: Math Scale Score Frequency Distribution Grade 5

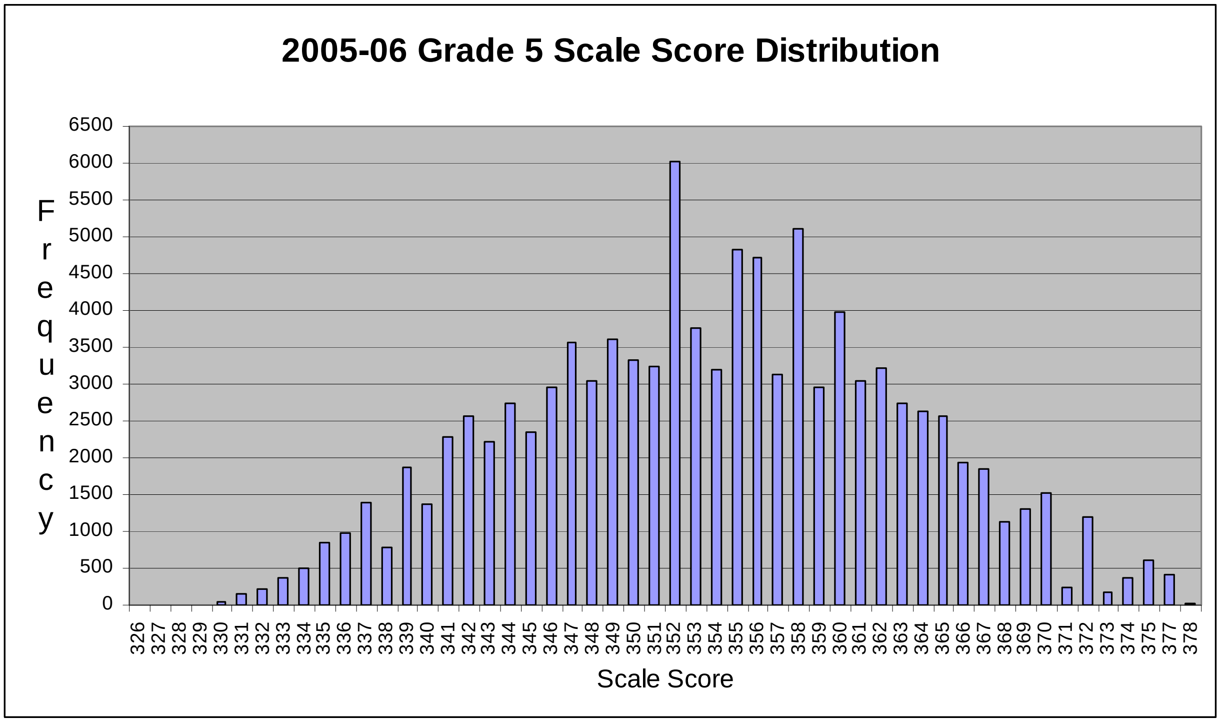


Figure 10: Math Scale Score Frequency Distribution Grade 6

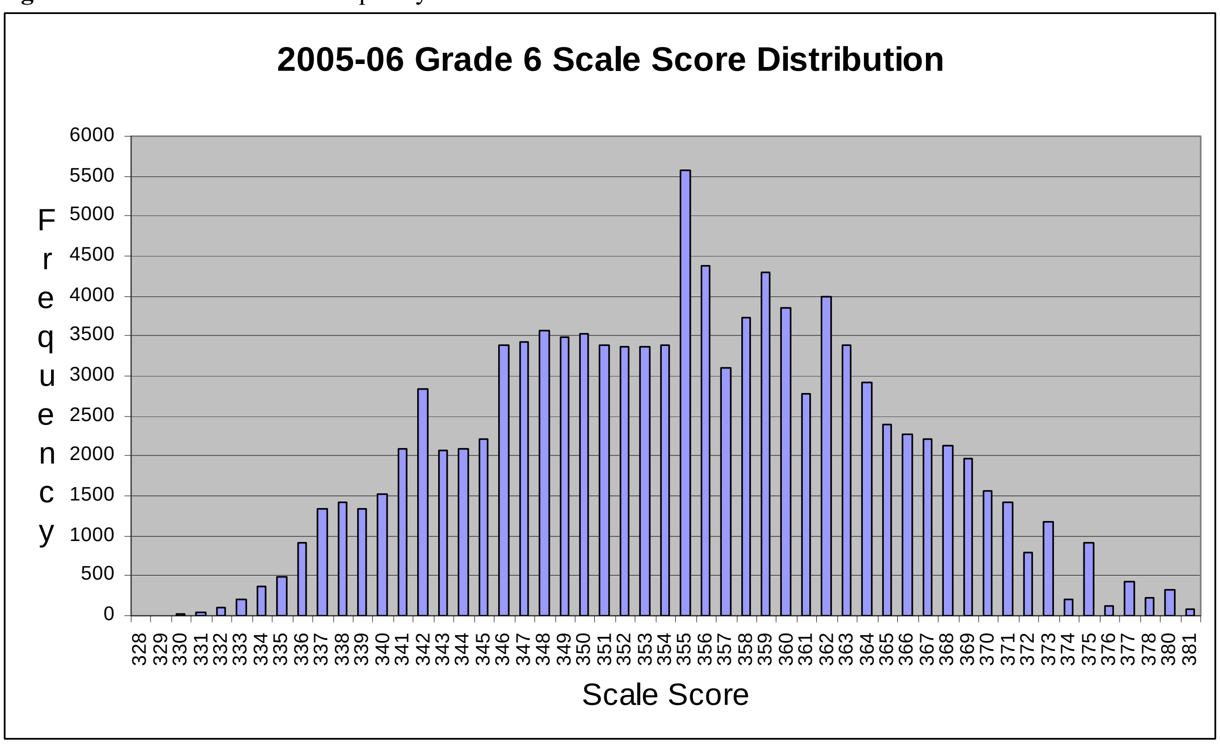


Figure 11: Math Scale Score Frequency Distribution Grade 7

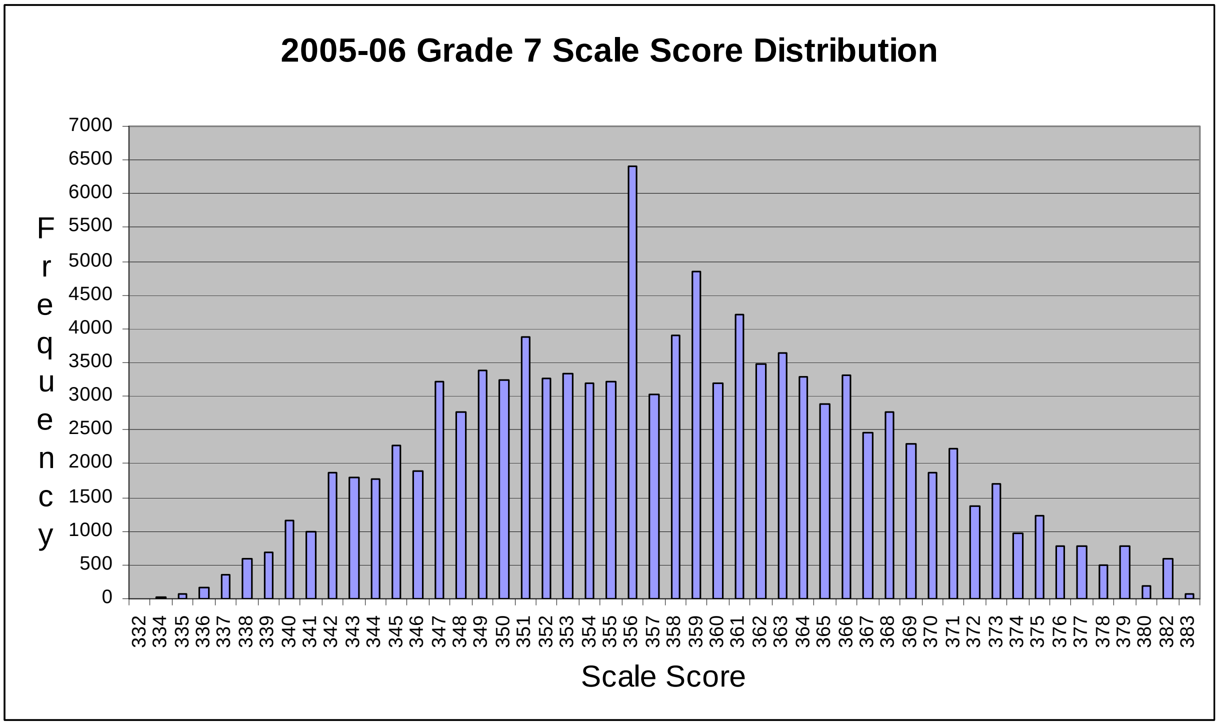


Figure 12: Math Scale Score Frequency Distribution Grade 8

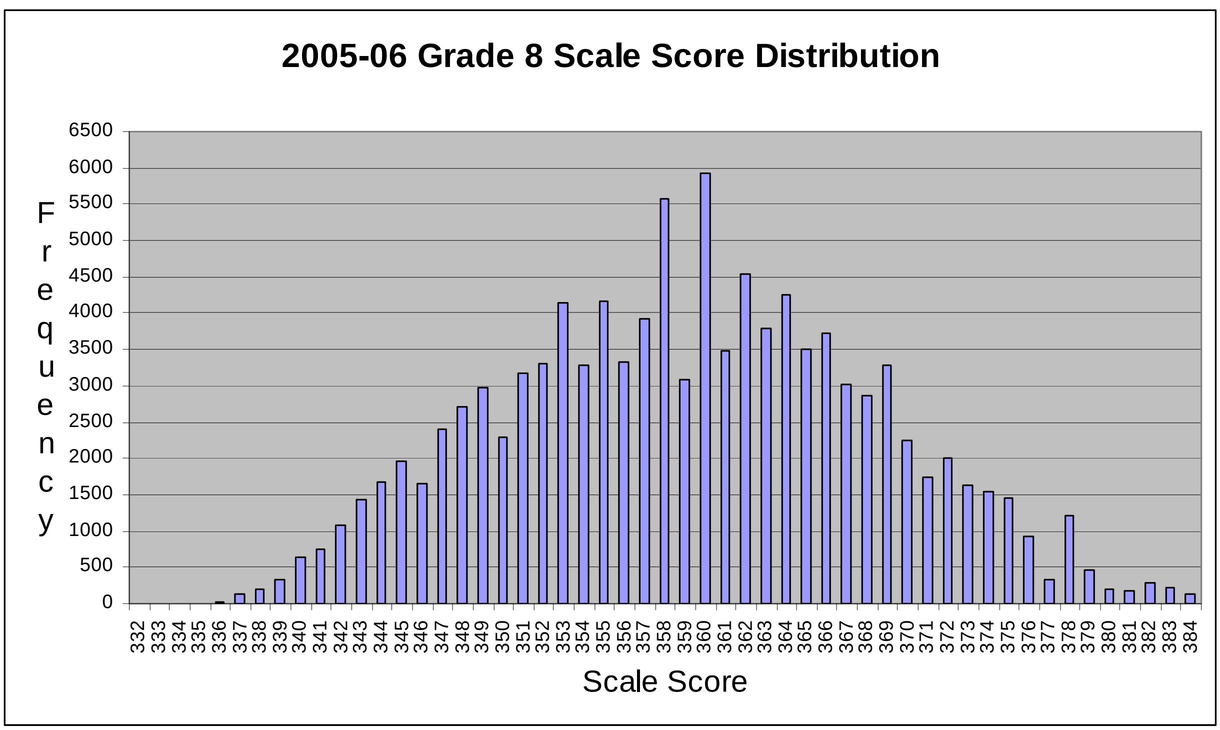


Figure 13: Algebra I Scale Score Frequency Distribution

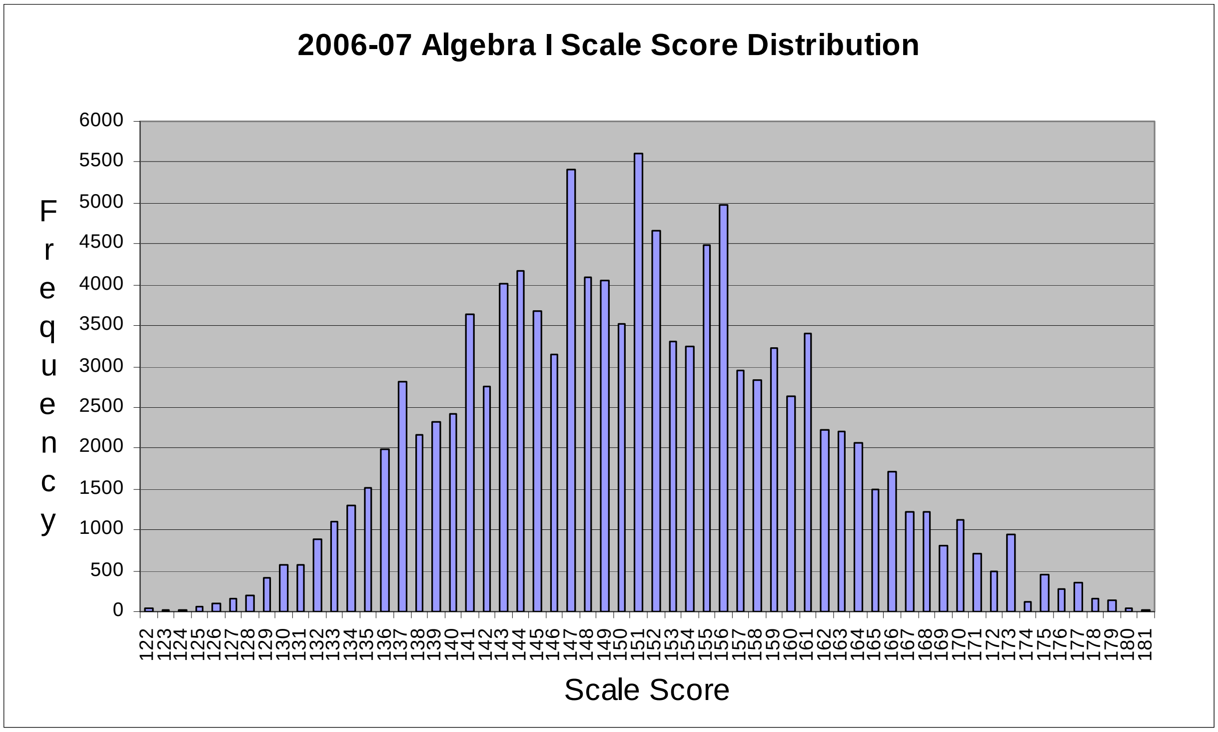


Figure 14: Geometry Scale Score Frequency Distribution

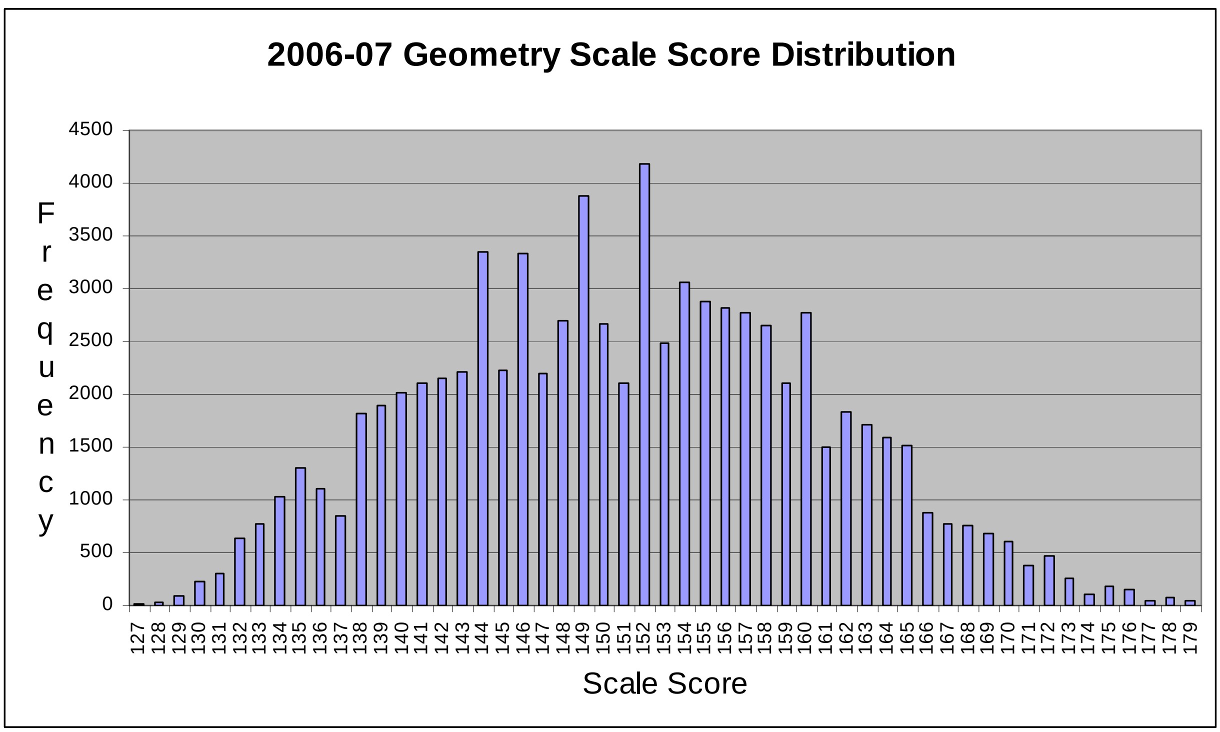
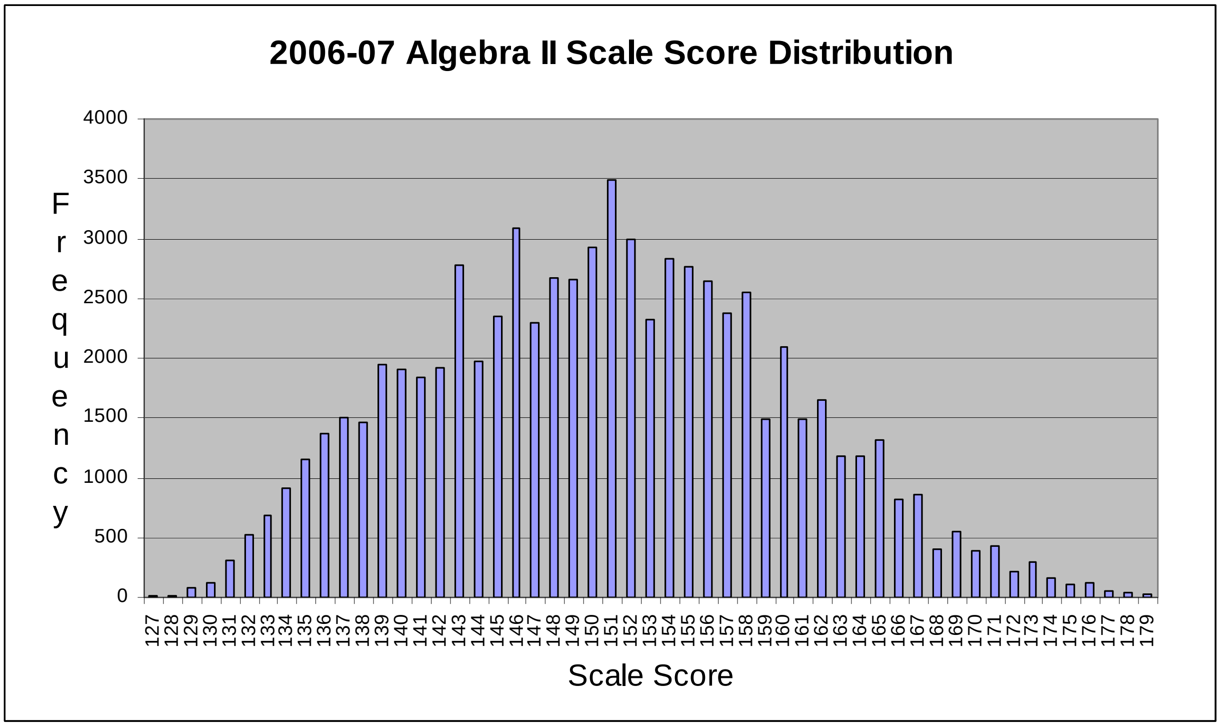


Figure 15: Algebra II Scale Score Frequency Distribution



6.5 Reliability of the North Carolina Mathematics Tests

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 are reliable and replicable. For a high-stakes multiple-choice test, the reliability coefficient should be 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 generally used to establish reliability for the North Carolina EOG and EOC Tests of Mathematics.

However, in the 2006–2007 school year, a special study was conducted with five of the NC EOC tests, among which was the Algebra I test. For the first time, the state offered, as an option, an online administration of the EOC tests. Given that the tests are being administered in a different mode, it was incumbent upon the state to demonstrate the comparability of the paper-and-pencil administration to the computer-delivered administration. As part of the research design, one of the data collections was a within-subjects design; that is, students took both a paper-and-pencil form and a computer-administered form. In order to maintain a high motivation condition, the student and district were allowed to use the higher of the two scores for purposes of grading and local accountability. Administration modes were, to the extent possible, counterbalanced, and no effort was made to control for students taking the same form of the test (for Algebra I, this happened for less than 450 students out of nearly 4,000 in the study). One of the outcomes of this study was to obtain an estimate of alternate forms and test-retest reliability.

Table 28: Reliability Coefficients from comparability study (Correlation between Scores from Paper-and-pencil and Computer-Based Administrations)

Subject – Pairing	Average Correlation	Range
Algebra I – alternate forms	0.898	0.844 – 0.931
Algebra I – test-retest	0.908	0.861 – 0.932

Of course, there may be some mode effect impacting the coefficients, but the results of the comparability study (see section 6.9 below) indicate that the forms are in fact comparable and mode effect is negligible. The perfect study would also have included paper-and-pencil replications and those coefficients would serve as a baseline.

6.6 Internal Consistency of the North Carolina Mathematics Tests

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

Table 29: Reliability indices averaged across North Carolina EOG and EOC Tests of Mathematics forms

Grade/Subject	N Operational Items on Test	Average Coefficient Alpha	Range of Coefficients Alpha
Pretest—Grade 3	38	0.888	0.876 – 0.894
Grade 3	50	0.905	0.901 – 0.914
Grade 4	50	0.915	0.911 – 0.919
Grade 5	50	0.913	0.905 – 0.919
Grade 6	50	0.915	0.909 – 0.919
Grade 7	50	0.922	0.917 – 0.929
Grade 8	60	0.920	0.914 – 0.925
Algebra I	64	0.920	0.916 – 0.925
Geometry	60	0.924	0.918 – 0.927
Algebra II (2006-07)	64	0.934	0.929 – 0.940
Algebra II (2007-08)[1]	51	0.909	0.897 – 0.915

[1]Based on the results of the 2006–2007 administration, it was determined that the Algebra II test could be shortened and still maintain adequate reliability. The 2007–2008 overall reliabilities for the new, shortened forms are included.

As noted above, the North Carolina EOG and EOC Tests of Mathematics are highly reliable as a whole. In addition, it is important to note that this high degree of reliability extends across gender, ethnicity, LEP status, migrant status, Title I status, and disability. Looking at coefficients alpha for the different groups reveals that across all test forms, in all grades and subjects, 57% of the values were at or above 0.90 and all but 5 (97% of all reliability coefficients) were above 0.85.

Table 30: Reliability indices averaged across North Carolina EOG and EOC Test of Mathematics forms (Gender)

Grade / Subject	Females	Males
Pretest—Grade 3	0.883	0.889
Grade 3	0.903	0.908
Grade 4	0.911	0.918
Grade 5	0.910	0.916
Grade 6	0.911	0.918
Grade 7	0.920	0.925
Grade 8	0.917	0.923
Algebra I	0.916	0.925
Geometry	0.923	0.925
Algebra II (64 items)	0.932	0.936
Algebra II (51 items)	0.906	0.912

Table 31: Reliability indices averaged across North Carolina EOG and EOC Test of Mathematics forms (Ethnicity)

Grade / Subject	Asian	Black	Hispanic	Native American	Multi-Racial	White
Pretest—Grade 3	0.892	0.850	0.847	0.859	0.878	0.887
Grade 3	0.913	0.875	0.881	0.872	0.894	0.899
Grade 4	0.929	0.877	0.892	0.889	0.910	0.911
Grade 5	0.926	0.877	0.894	0.896	0.908	0.910
Grade 6	0.932	0.867	0.890	0.875	0.908	0.912
Grade 7	0.933	0.876	0.905	0.890	0.915	0.921
Grade 8	0.935	0.867	0.896	0.882	0.913	0.921
Algebra I	0.935	0.882	0.908	0.894	0.914	0.918
Geometry	0.941	0.878	0.910	0.895	0.918	0.921
Algebra II (64 items)	0.951	0.896	0.919	0.893	0.931	0.933
Algebra II (51 items)	0.928	0.857	0.896	0.838	0.894	0.910

Table 32: Reliability indices averaged across North Carolina EOG and EOC Tests of Mathematics forms (Other Characteristics)

Grade / Subject	No Disability	Disability	Not LEP	LEP	Not Title I	Title I	Not Migrant	Migrant
Pretest—Grade 3	0.887	0.885	0.887	0.791	0.890	0.876	0.889	0.839
Grade 3	0.902	0.900	0.906	0.861	0.906	0.897	0.905	0.867
Grade 4	0.913	0.897	0.915	0.869	0.917	0.906	0.915	0.885
Grade 5	0.910	0.891	0.913	0.880	0.915	0.905	0.913	0.895
Grade 6	0.913	0.879	0.915	0.855	0.917	0.898	0.915	0.870
Grade 7	0.920	0.890	0.922	0.870	0.924	0.907	0.922	0.871
Grade 8	0.918	0.876	0.920	0.858	0.922	0.903	0.920	0.893
Algebra I	0.919	0.884	0.920	0.896	0.920	0.927	0.921	0.894
Geometry	0.924	0.895	0.924	0.905	0.924	0.928	0.924	0.879
Algebra II (64 items)	0.934	0.910	0.934	0.927	0.934	0.882	0.934	0.894
Algebra II (51 items)	0.888	0.856	0.909	0.879	0.909	0.852	0.909	0.814

There was some variation among forms. Coefficients alpha that were below the 0.85 threshold were

Grade 3, Form P for the migrant subgroup (0.845)
Grade 6, form K for the Native American subgroup (0.846)
Grade 6, forms K and P for the migrant subgroup (0.832 and 0.849, respectively)
Grade 6, forms K and L for the LEP subgroup (0.843 and 0.845, respectively)
Grade 8, form M for the LEP subgroup (0.846)
Algebra II, form K for the Black / African American subgroup (0.829)
Algebra II, forms K and L for the Native American subgroup (0.781 and 0.823, respectively)
Algebra II, form K for students with disabilities (0.829)
Algebra II, form K for the LEP subgroup (0.844)
Algebra II, forms K and N for the Title I subgroup (0.836 and 0.842, respectively)
Algebra II, forms K, L, and M for the migrant subgroup (0.769, 0.795, and 0.834, respectively)

The Pretest—Grade 3 presents some distinct challenges. All students take this assessment, as there currently are no alternate or alternative assessments available (they are under development). This test is also approximately ¾ the length of the other EOG math assessments. Although the overall reliability is still acceptable, there are some forms that are less reliable for certain subgroups, generally the smaller subgroups (i.e., migrant, LEP, Native American). In nearly all cases, when the test is prophesied to be the same length as the Grade 3 EOG Math Test, the reliability would exceed the 0.85 criterion. The exceptions are form P for migrant students (0.801, would prophesy to 0.841), and all forms for LEP students (range 0.762 to 0.809, would prophesy to 0.808 to 0.848).

The revised, shortened forms of Algebra II, as would be expected, have lower reliabilities than their longer counterparts, and for some of the very small subgroups, the shorter forms have reliabilities below the 0.85 threshold. Although in every other respect, form K of the Algebra II test behaves identically to the other three forms; this form has consistently lower reliabilities for many subgroups and will be the first form to be replaced.

Although the North Carolina Testing Program administers alternate forms of the test, it is not generally 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 in grades 3, 5, and 8 or Algebra I 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. A 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 ranges for scores on the North Carolina EOC and EOG Tests of Mathematics are provided in table 33 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 EOG Tests of Mathematics 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 generally around 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 5 to 6 points. This is typical as scores become more extreme due to less measurement precision associated with those extreme scores.

Table 33: Ranges of standard error of measurement for scale scores by grade or subject

Grade/Subject	SEM Range	SEM at I-II Cut Score	SEM at II-III Cut Score	SEM at III-IV Cut Score
Pretest—Grade 3	3–7	6	4	4
Grade 3	2–6	5	3	3
Grade 4	2–6	5	3	2
Grade 5	2–6	5	3	2
Grade 6	2–6	5	3	2
Grade 7	2–6	5	3	2
Grade 8	2–6	5	3	2
Algebra I	3–5	3	3	3
Geometry	2–5	5	3	2
Algebra II	2–5	4	3	2

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 16: Standard Errors of Measurement on the Pretest—Grade 3 of Mathematics Test forms

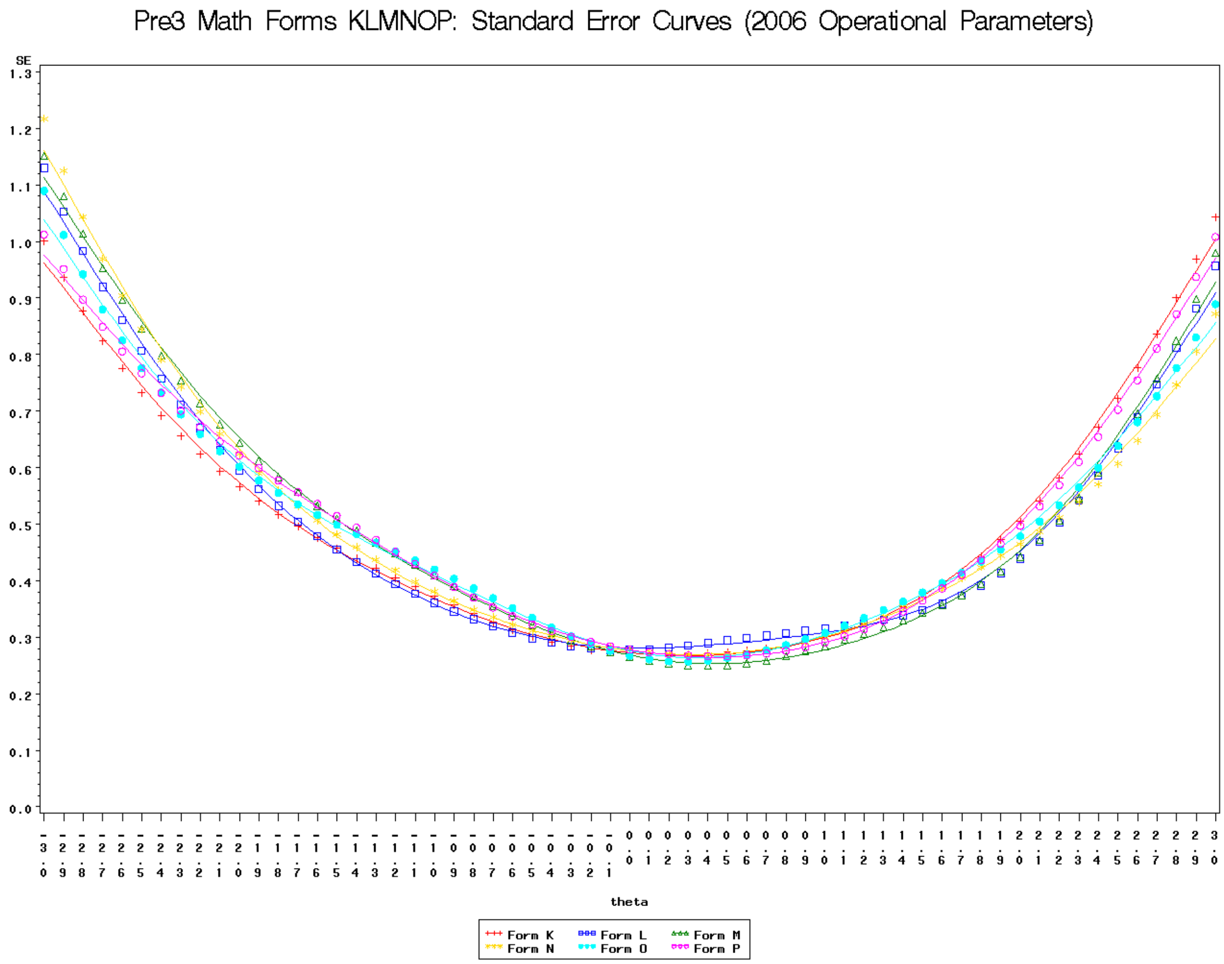


Figure 17: Standard Errors of Measurement on the Grade 3 Mathematics Test forms

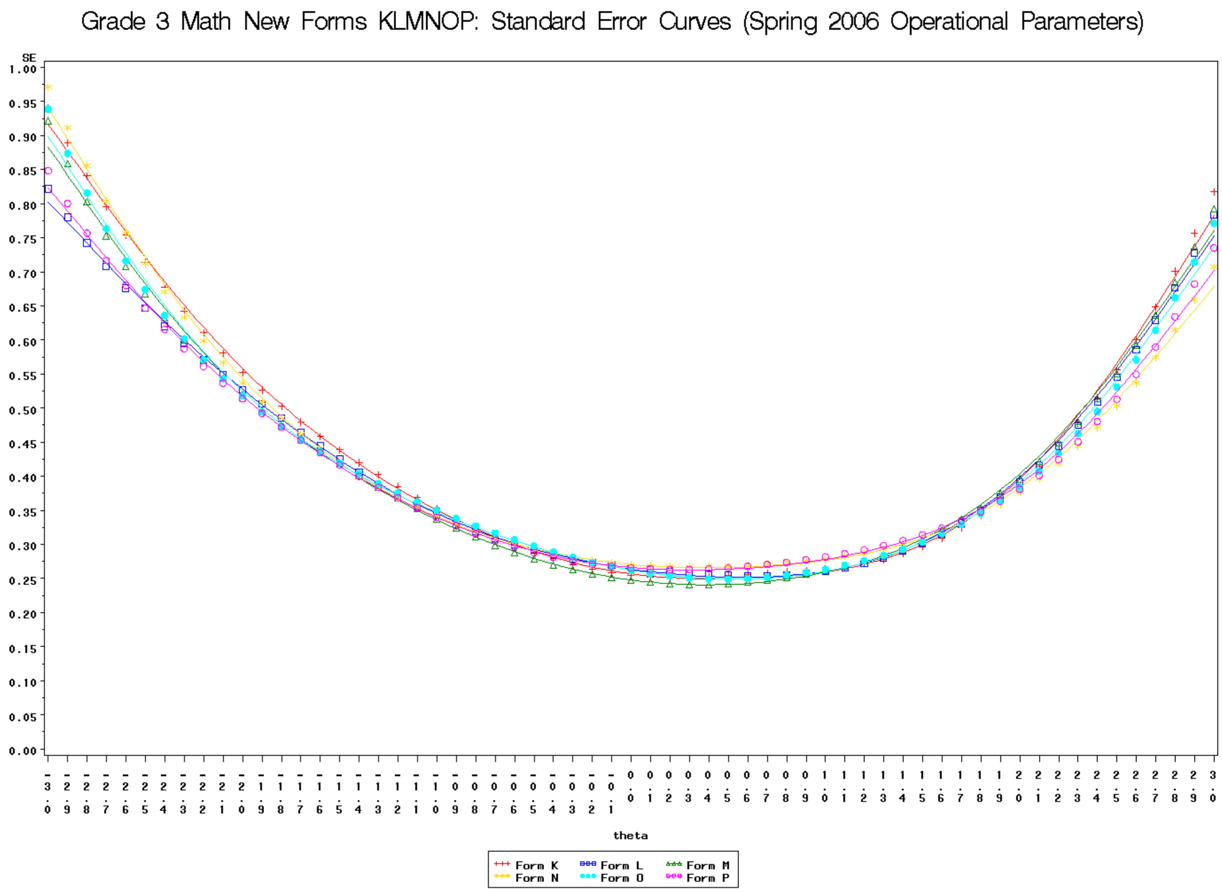


Figure 18: Standard Errors of Measurement on the Grade 4 Mathematics Test forms

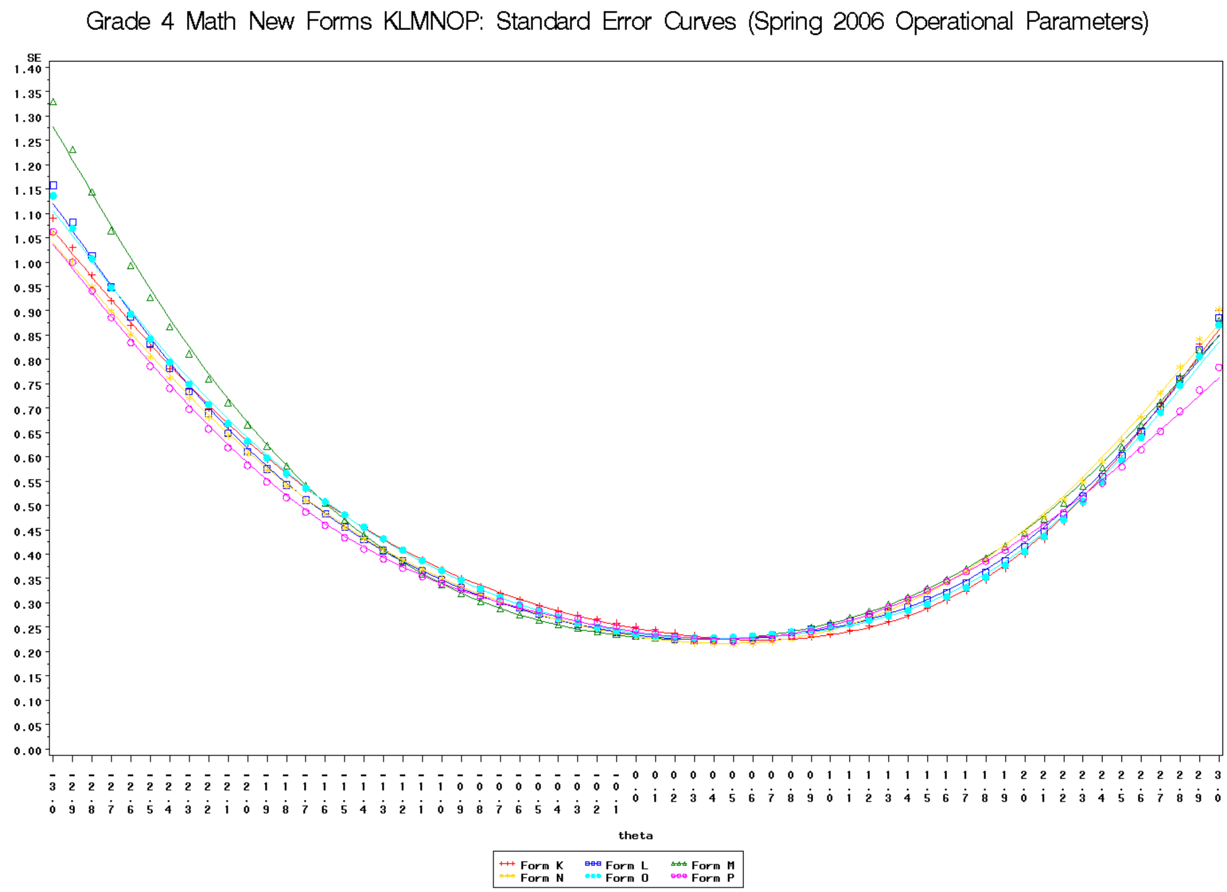


Figure 19: Standard Errors of Measurement on the Grade 5 Mathematics Test forms

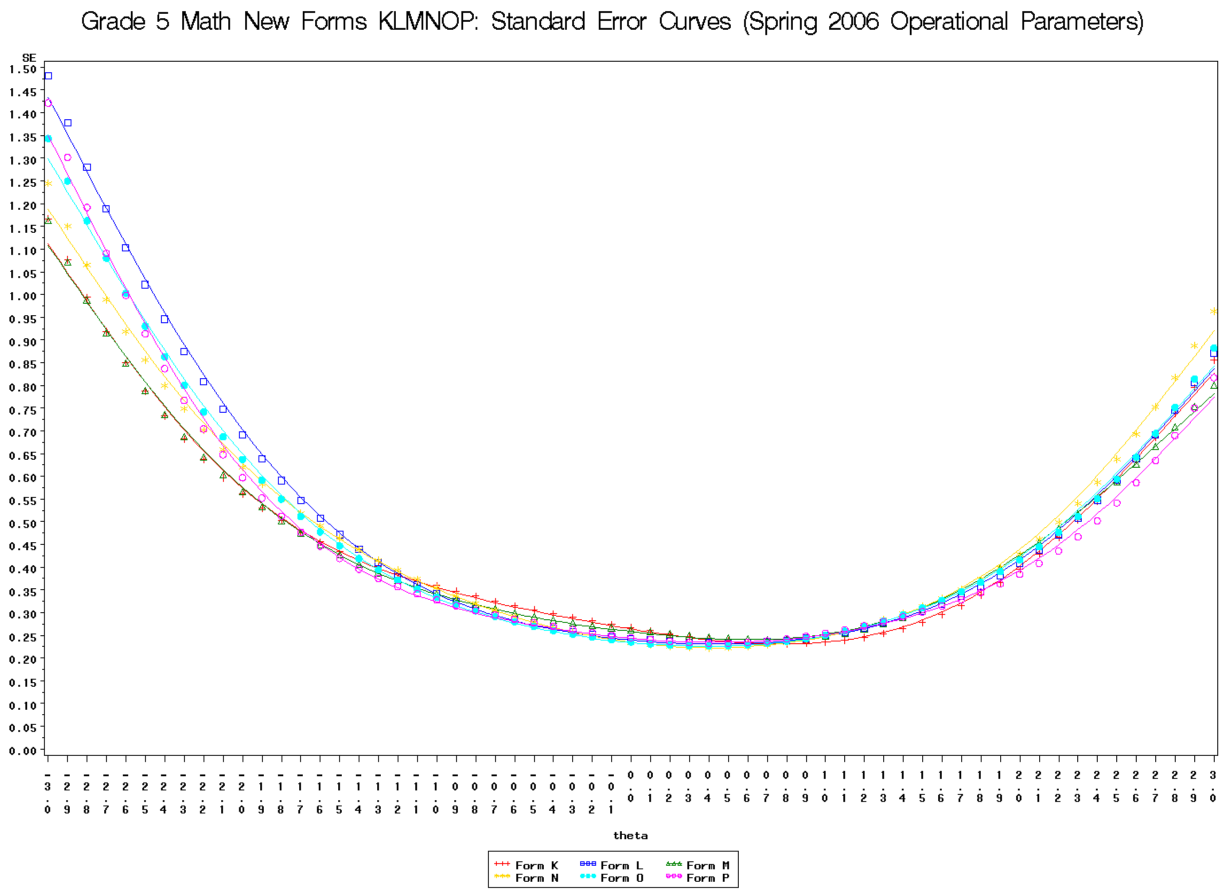


Figure 20: Standard Errors of Measurement on the Grade 6 Mathematics Test forms

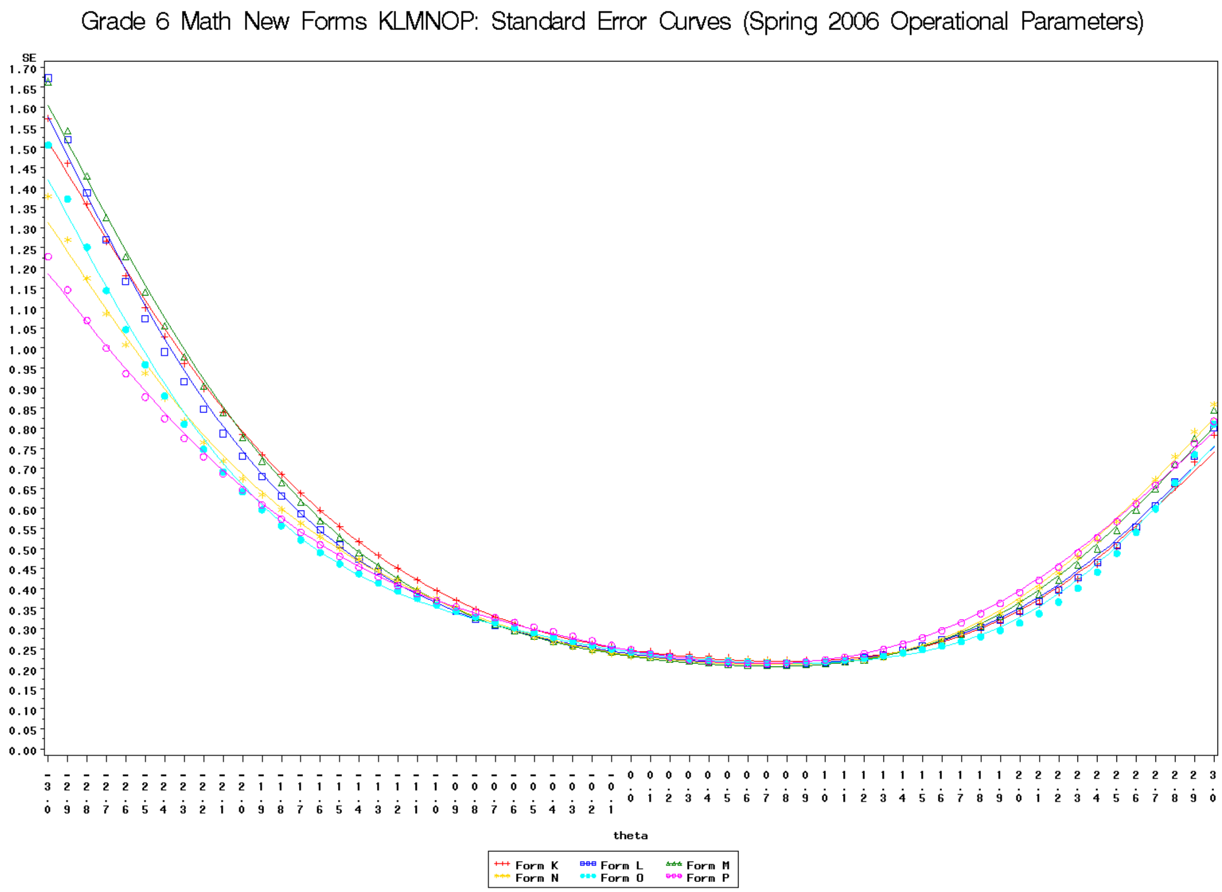


Figure 21: Standard Errors of Measurement on the Grade 7 Mathematics Test forms

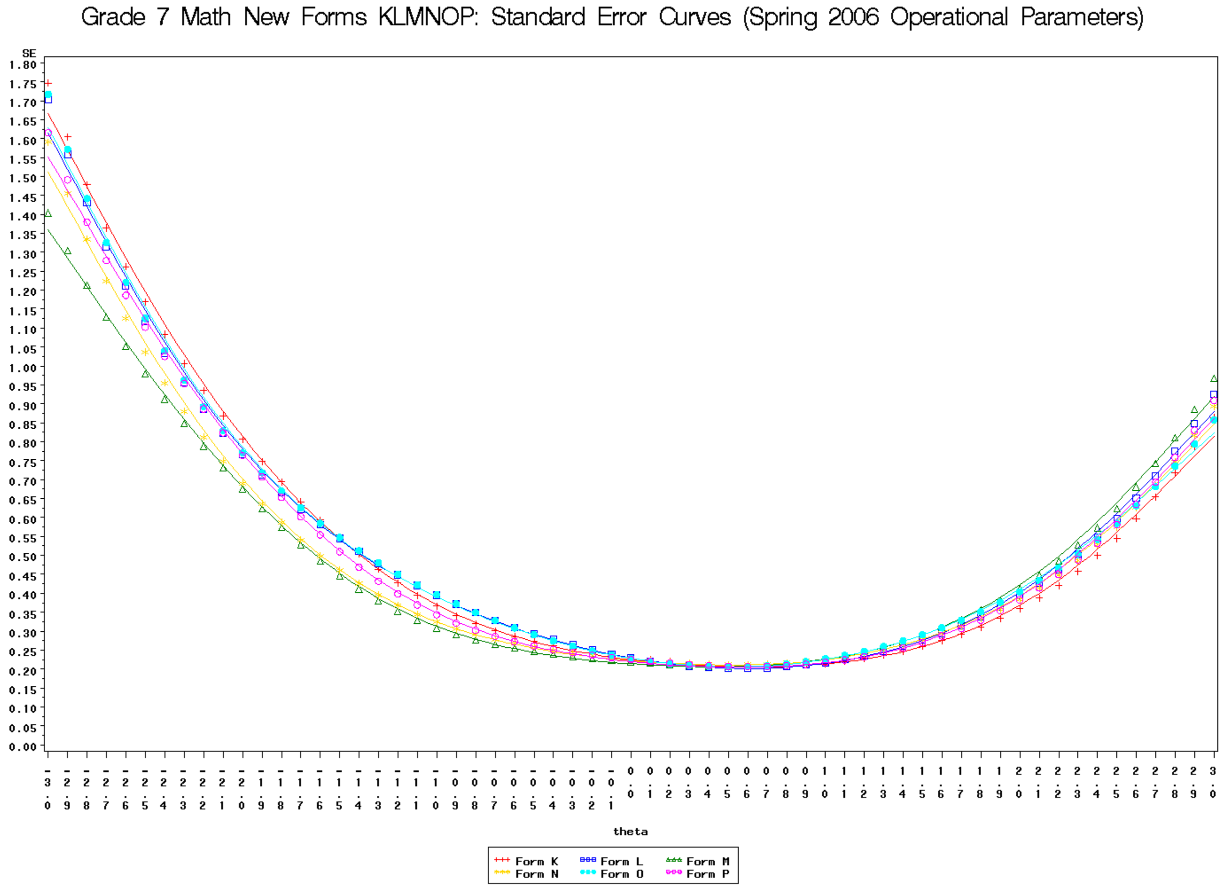


Figure 22: Standard Errors of Measurement on the Grade 8 Mathematics Test forms

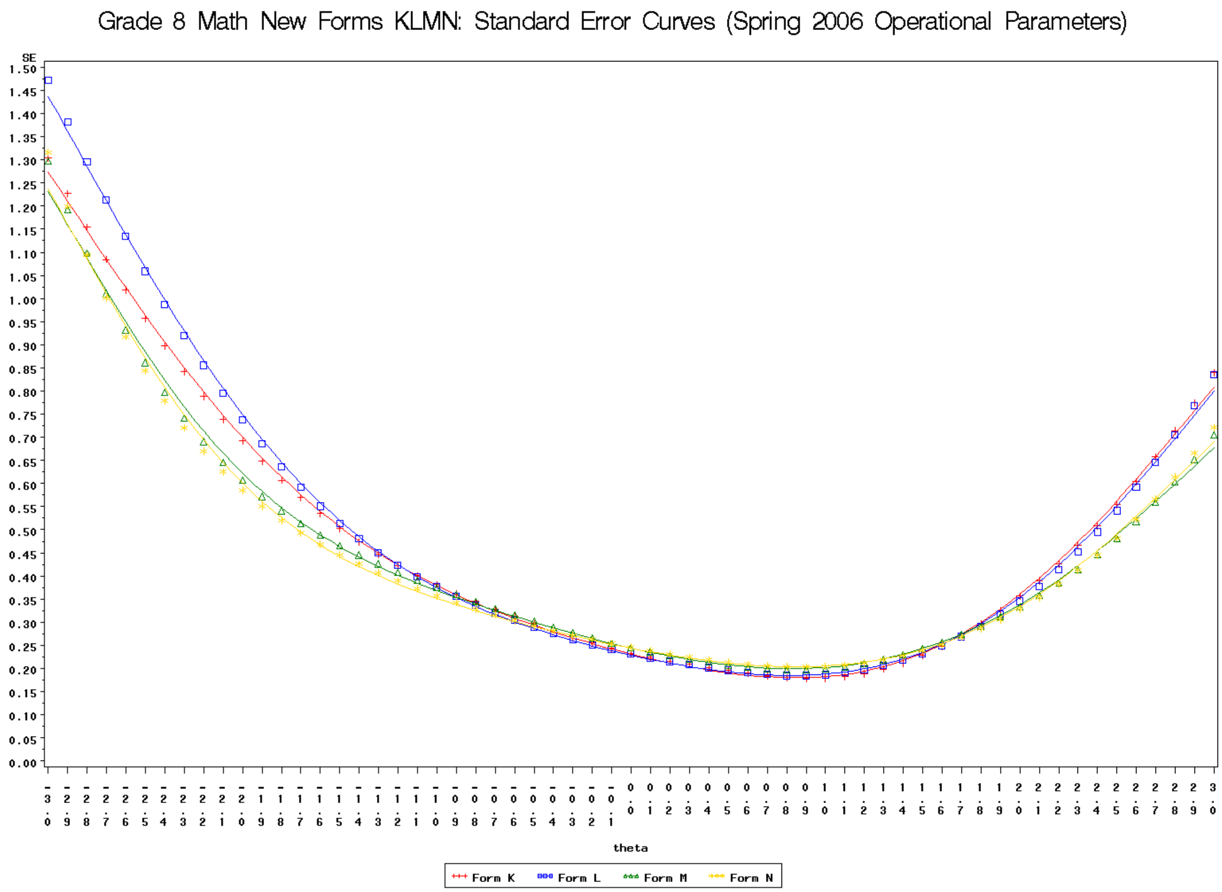


Figure 23: Standard Errors of Measurement on the Algebra I Test forms

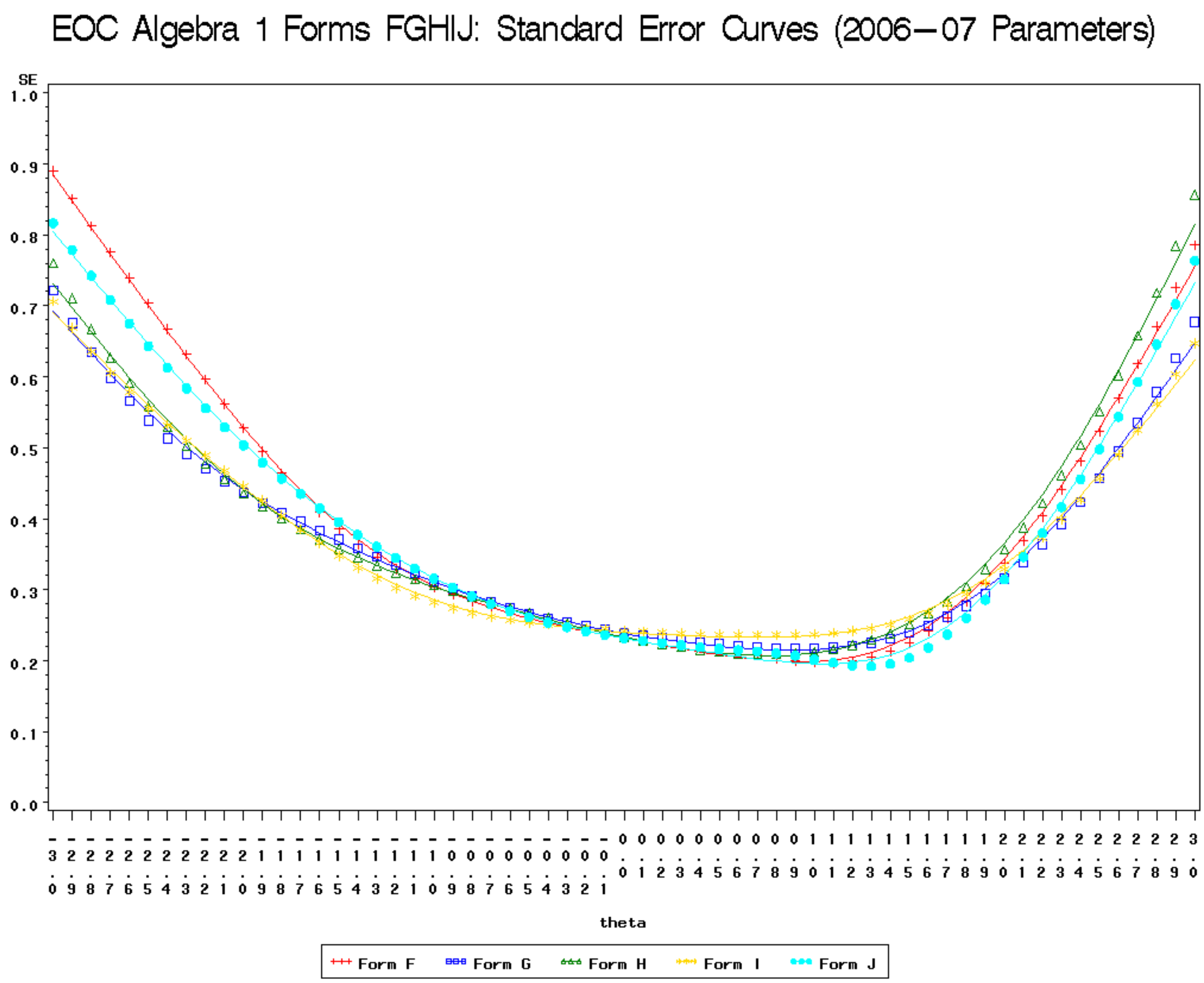


Figure 24: Standard Errors of Measurement on the Geometry Test forms

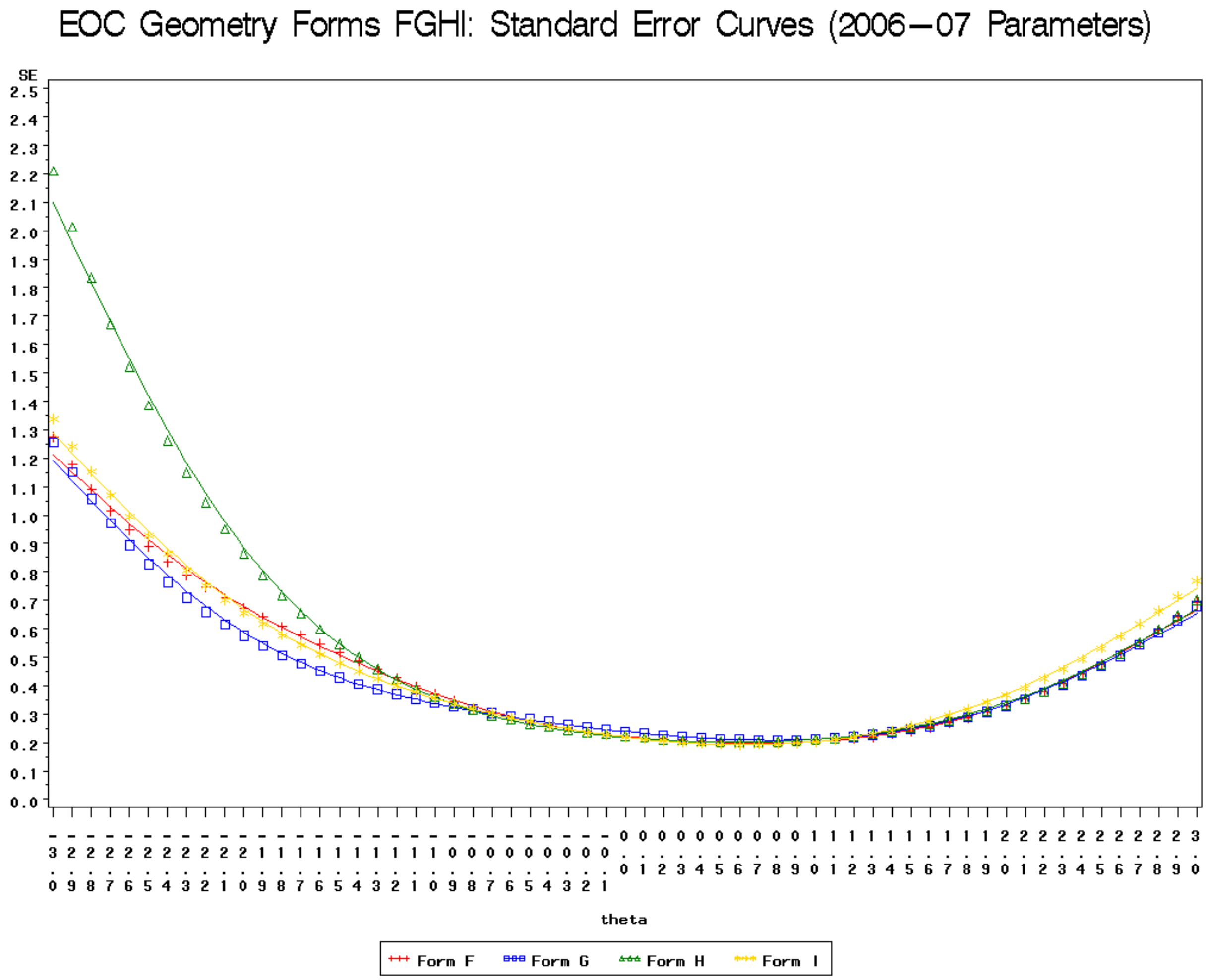


Figure 25: Standard Errors of Measurement on the Algebra II Test forms

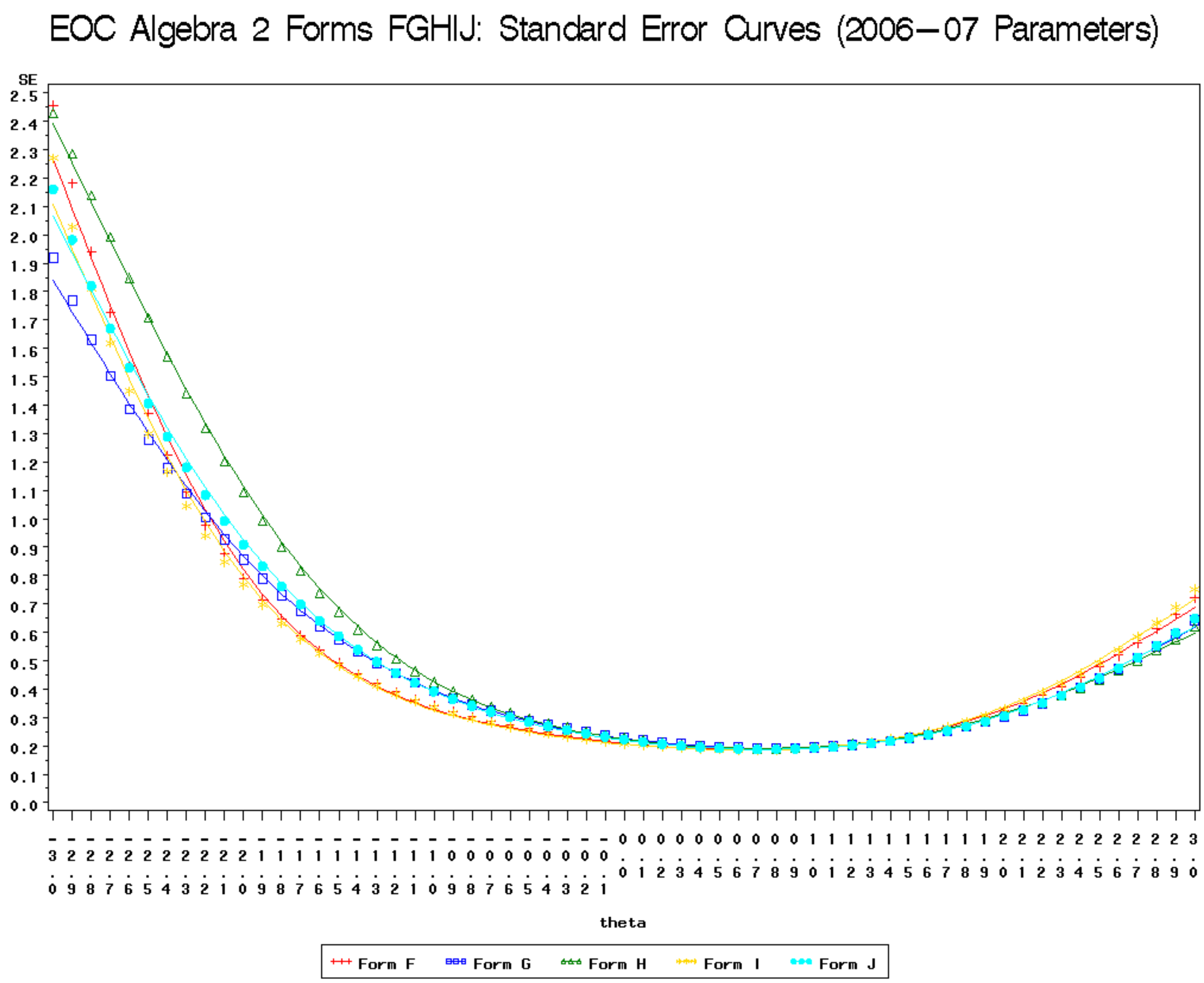
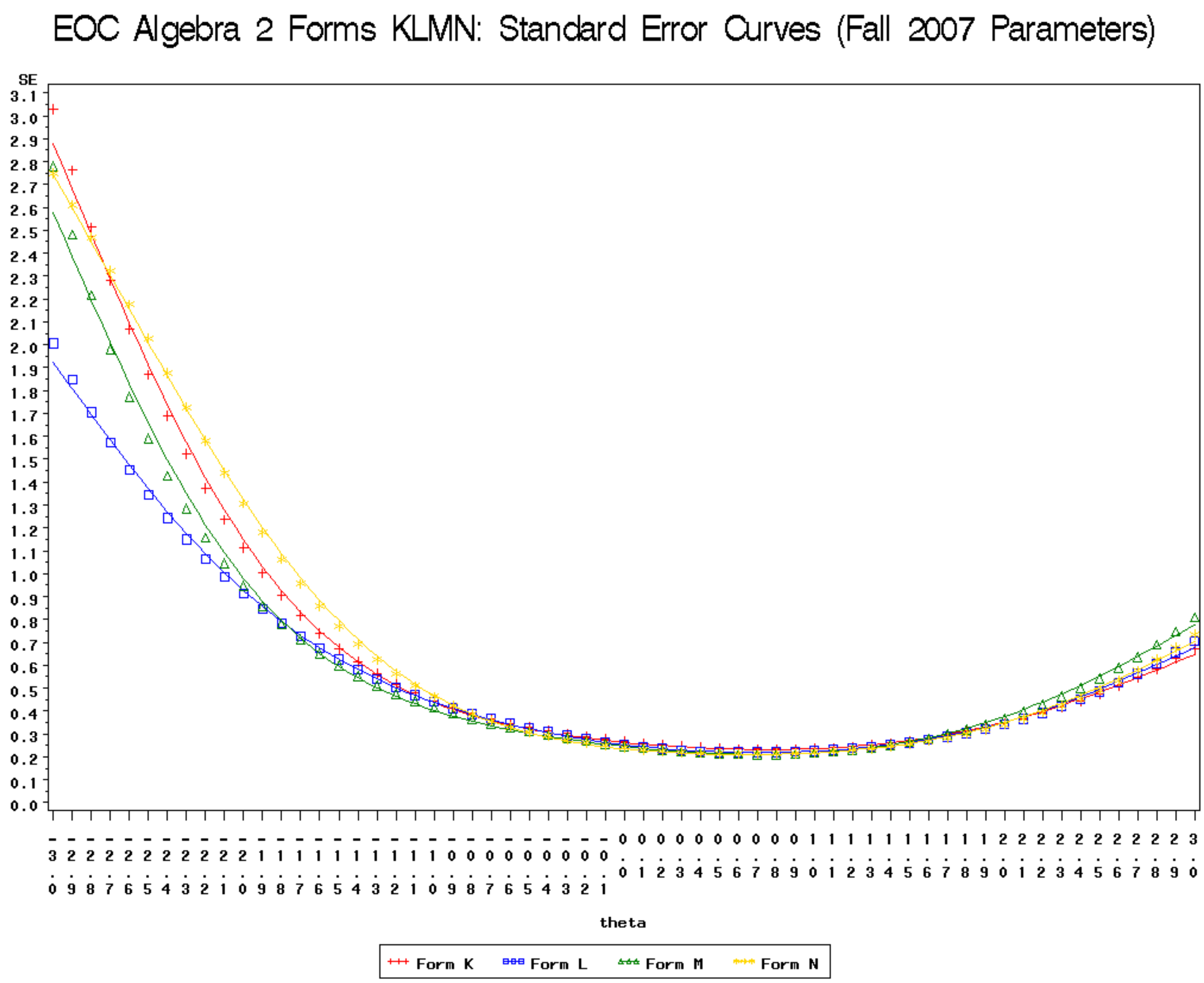


Figure 26: Standard Errors of Measurement on the revised Algebra II Test forms



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 at the goal level. That is, each form has the same number of items from the number and operations strand (Goal 1) as every other form administered in that grade. The specific questions asked on each form are a random domain sample of the topics in that grade'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. Additionally, the two parts of the mathematics tests, Calculator Active and Calculator Inactive, are also equivalent at the whole-score level. That is, all the Calculator Active portions of the tests for a given grade are equally difficult. However, due to the purposively random selection of items tested in each goal, the tests are not statistically equated at the goal level.

The use of multiple equivalent and parallel forms has given rise to several “urban legends,” foremost among which is that “The red form is harder” (referring to the color of the front cover of one of the base form test booklets). However, as the following figures show, the tests are indeed equivalent.

Figure 27: Test Characteristic Curves for the Pretest—Grade 3 of Mathematics Test forms

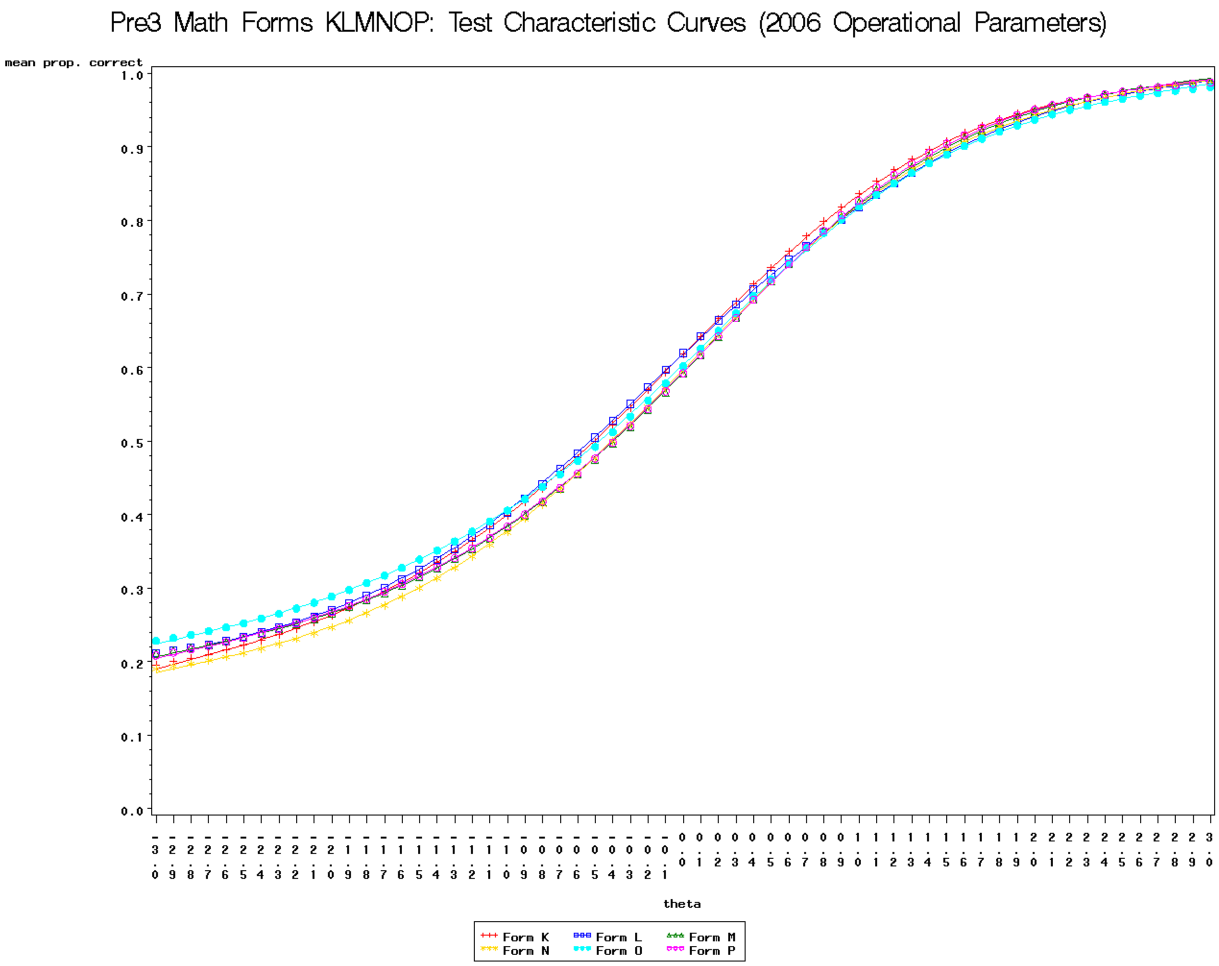


Figure 28: Test Characteristic Curves for the Grade 3 Mathematics Test forms

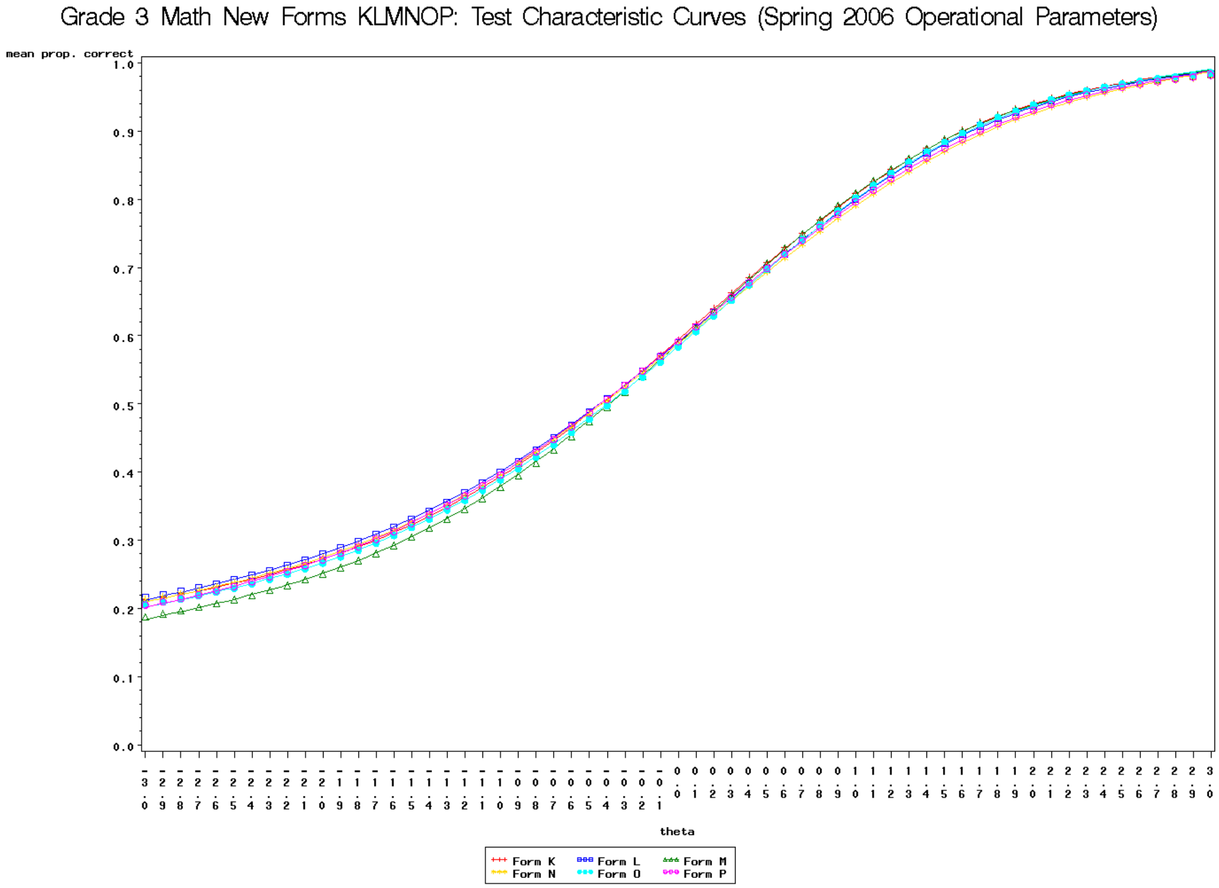


Figure 29: Test Characteristic Curves for the Grade 4 Mathematics Test forms

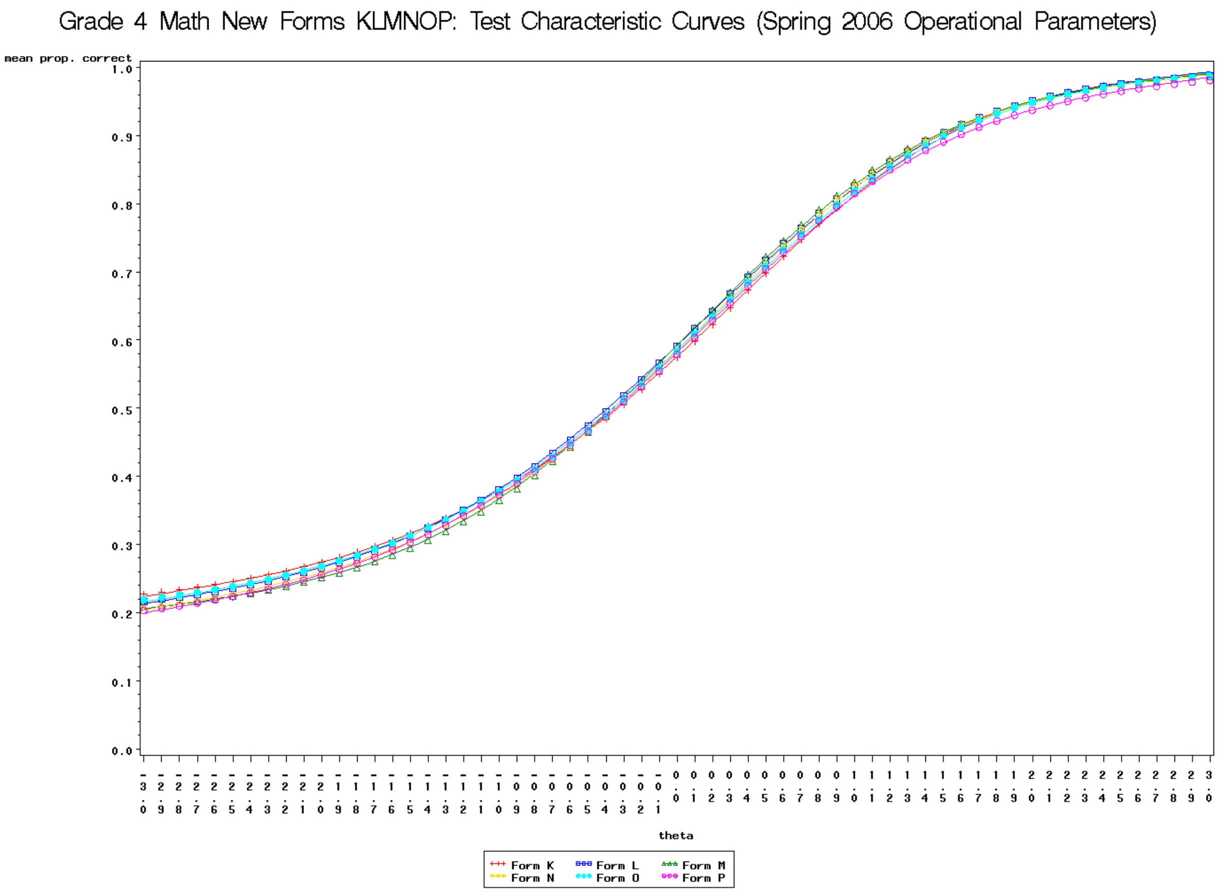


Figure 30: Test Characteristic Curves for the Grade 5 Mathematics Test forms

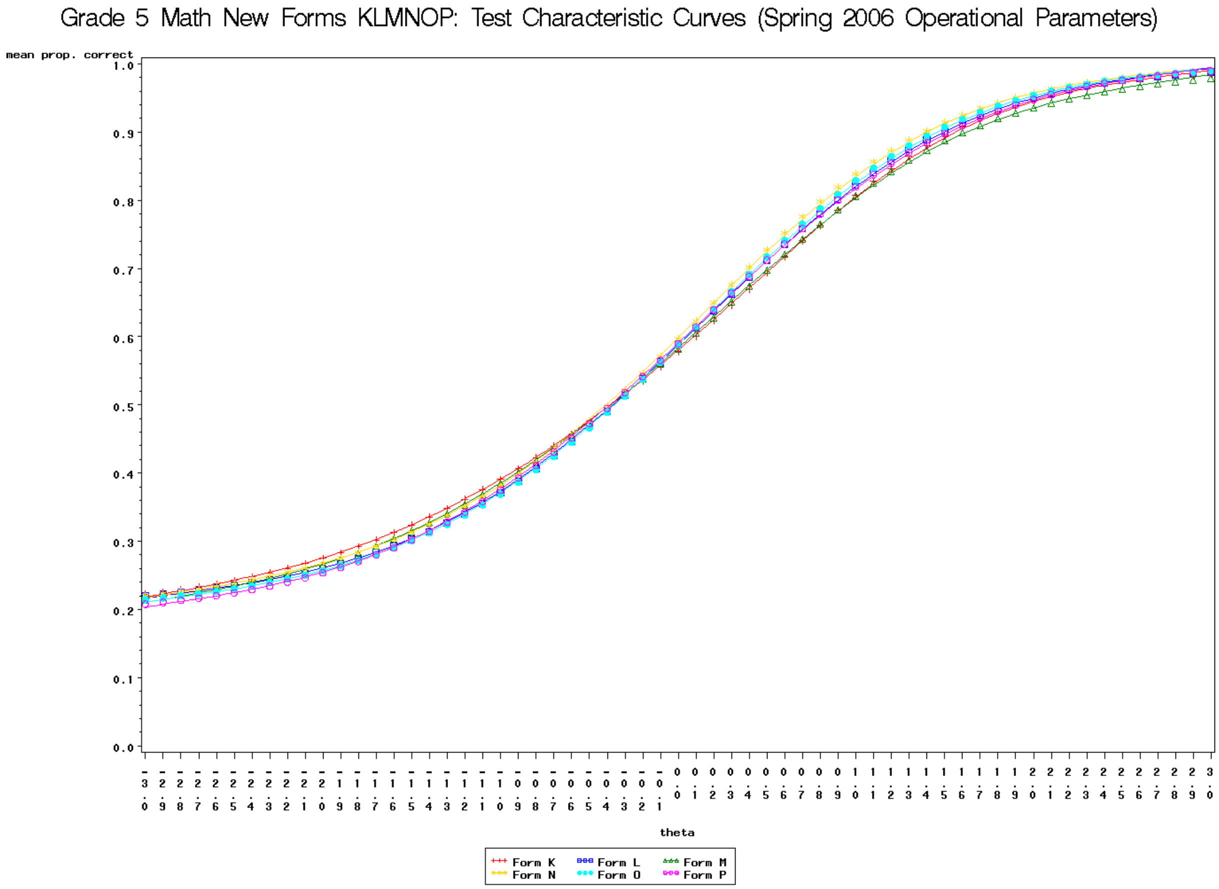


Figure 31: Test Characteristic Curves for the Grade 6 Mathematics Test forms

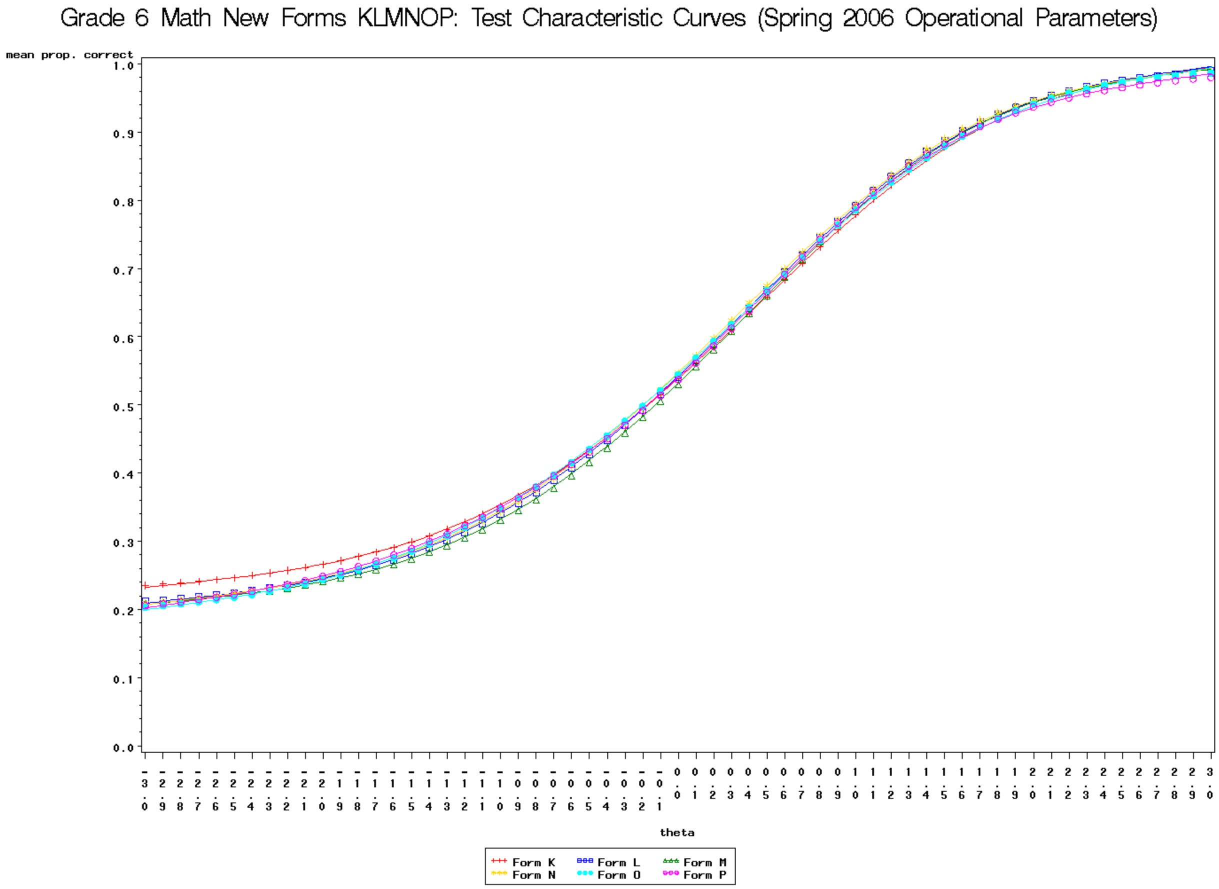


Figure 32: Test Characteristic Curves for the Grade 7 Mathematics Test forms

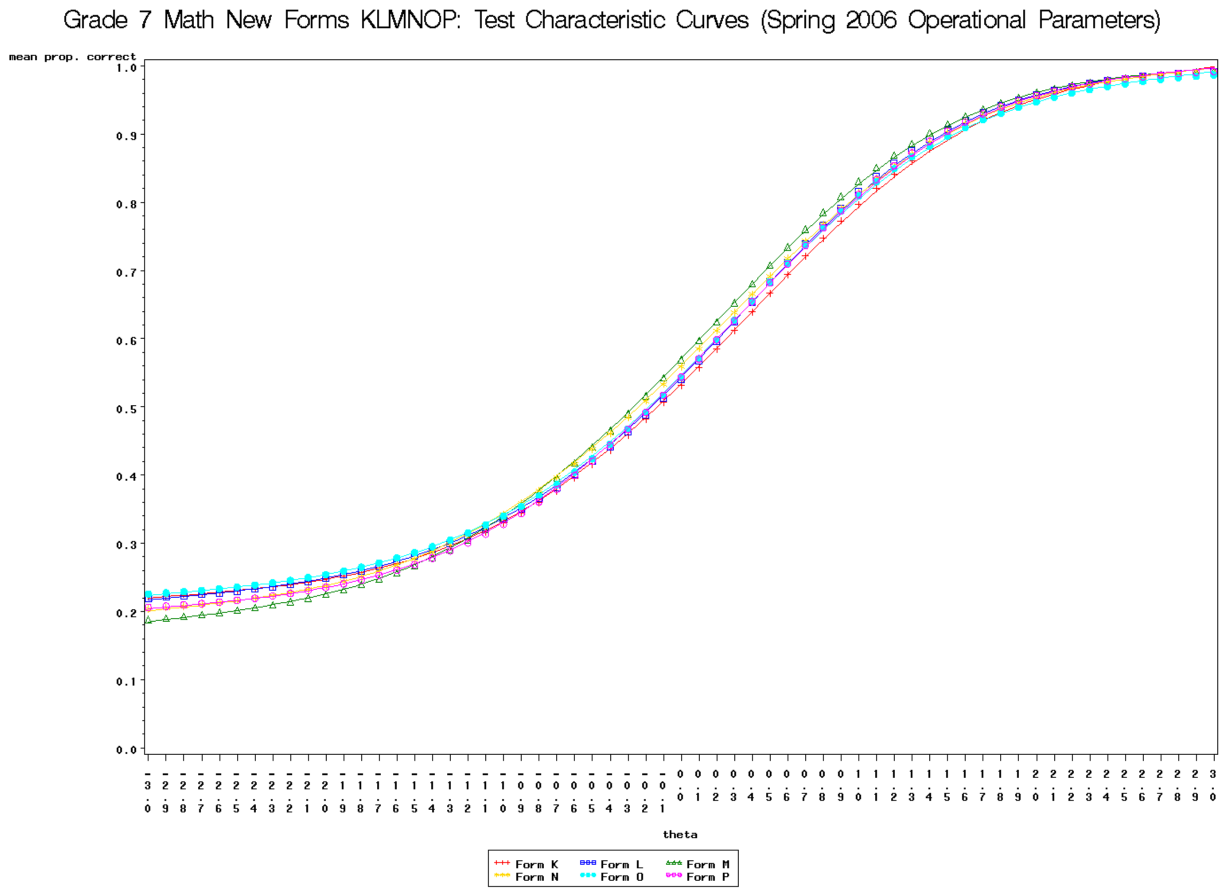


Figure 33: Test Characteristic Curves for the Grade 8 Mathematics Test forms

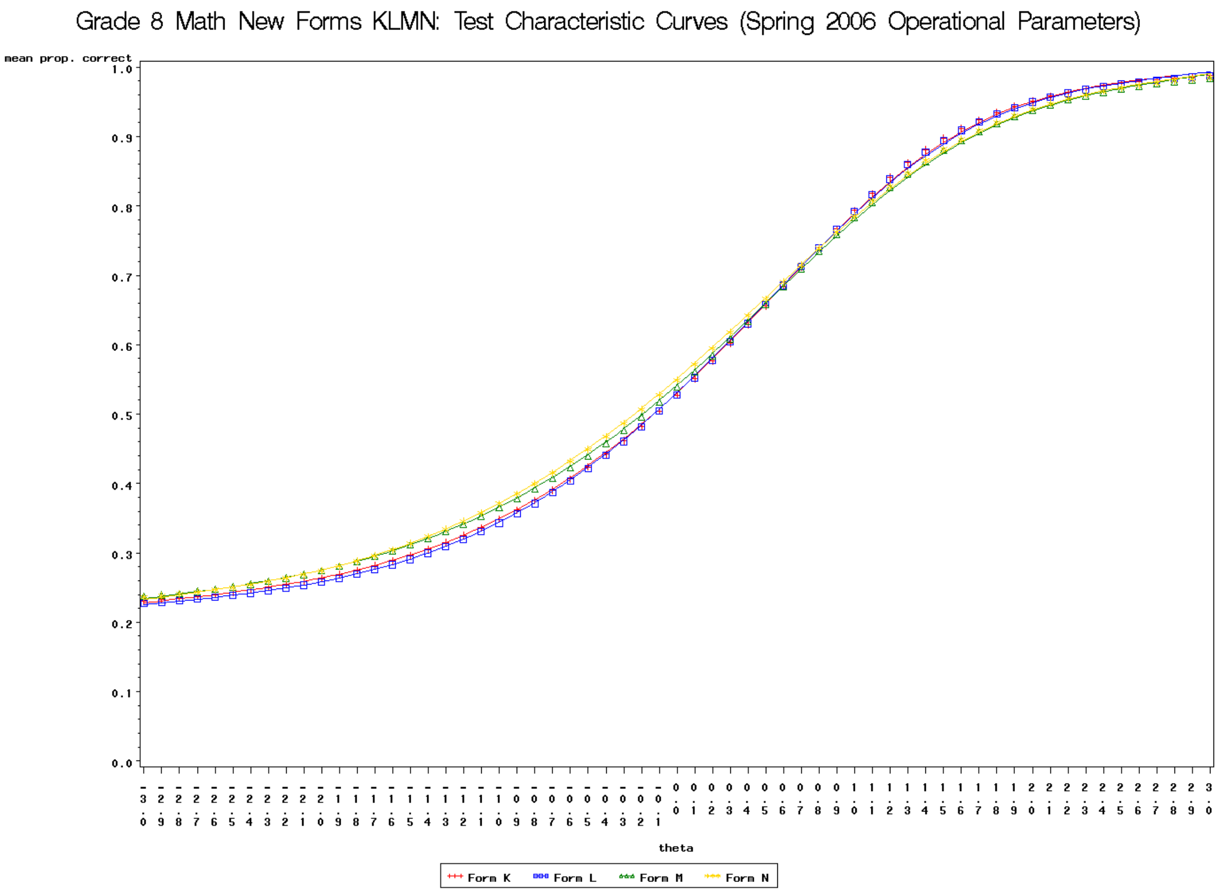


Figure 34: Test Characteristic Curves for the Algebra I Test forms

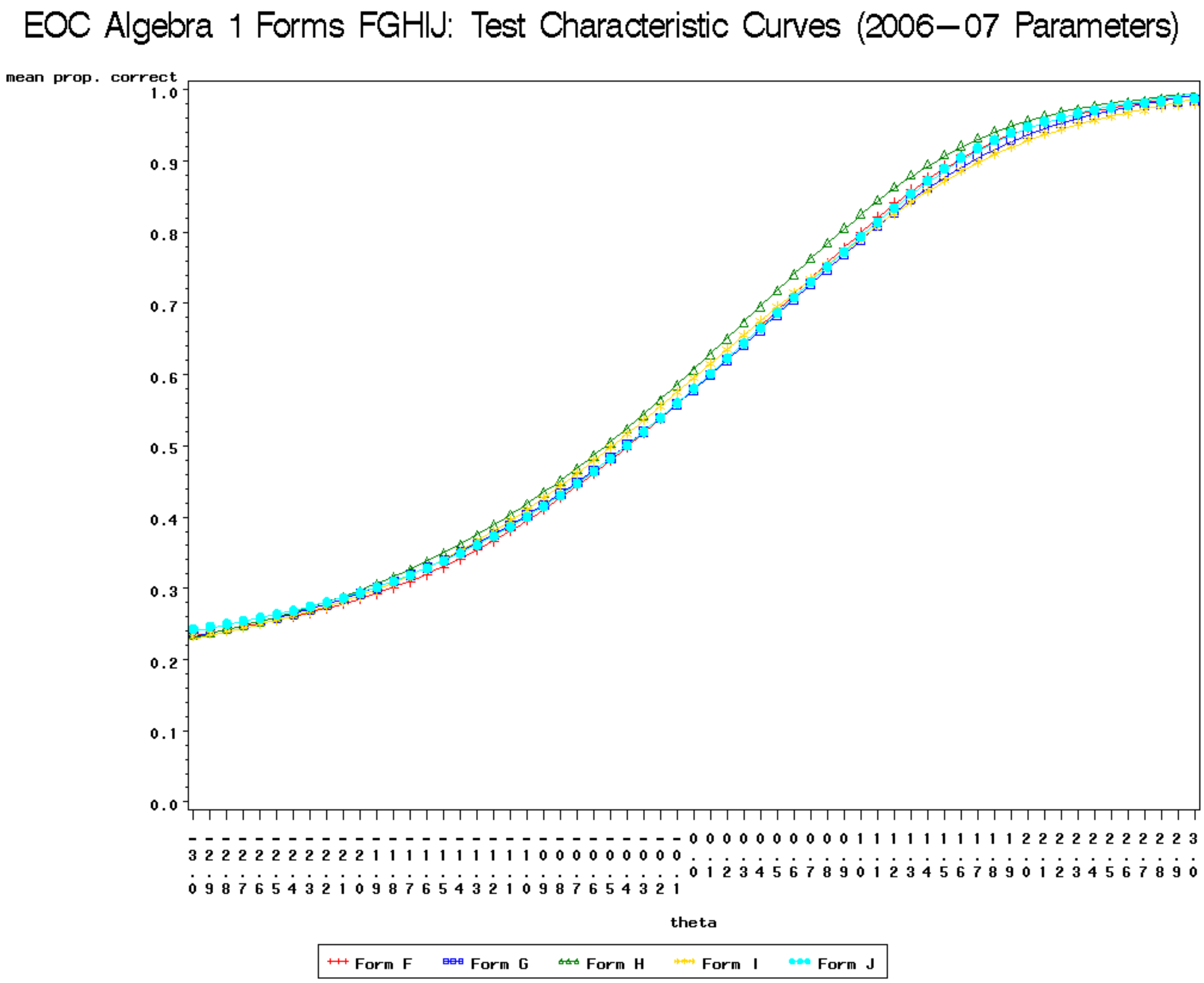


Figure 35: Test Characteristic Curves for the Geometry Test forms

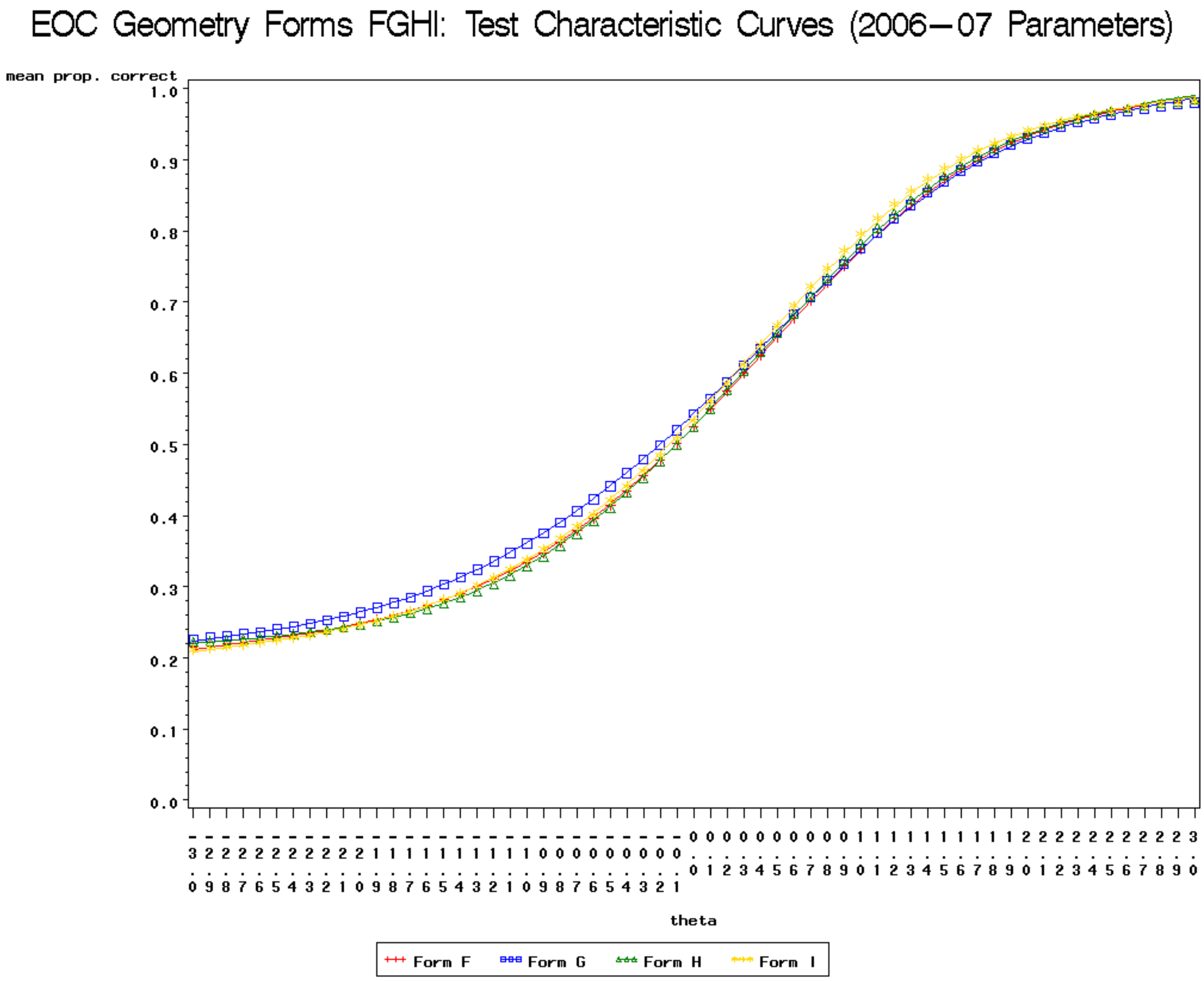


Figure 36: Test Characteristic Curves for the Algebra II Test forms

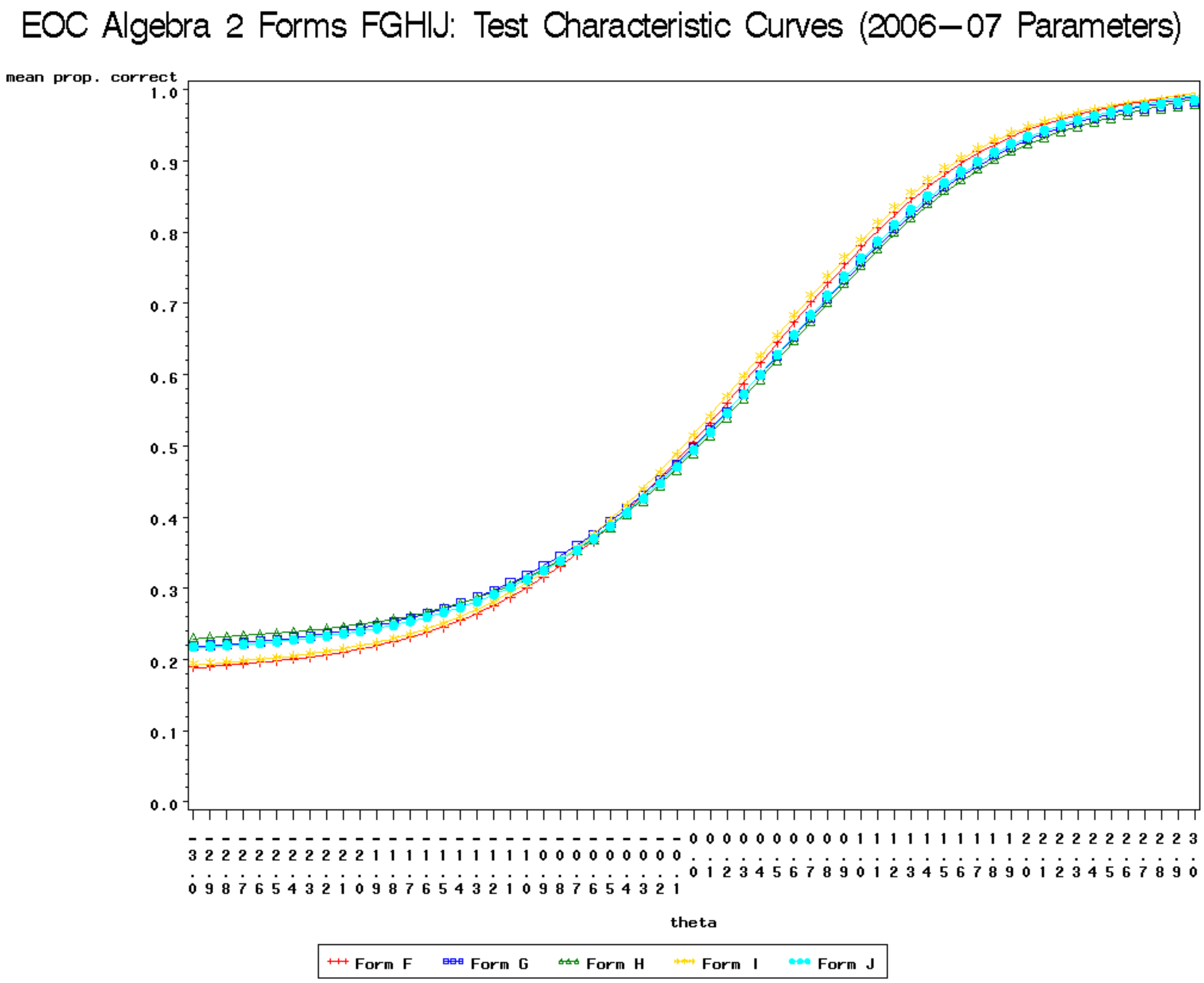
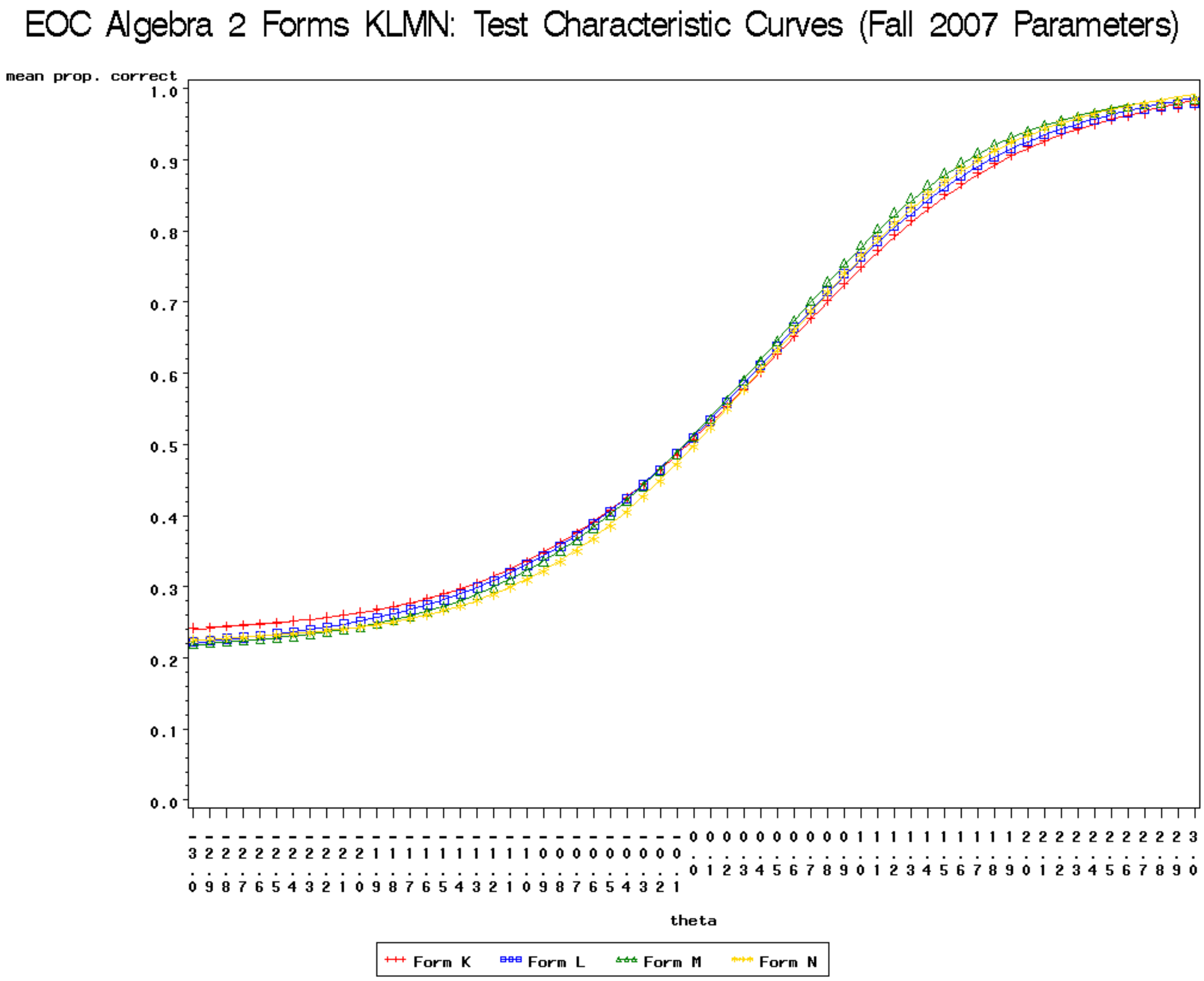


Figure 37: Test Characteristic Curves for the revised Algebra II Test forms



For each grade's 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.

6.9 Administration Mode Comparability

As previously mentioned, during the spring of 2007, a comparability study was designed to empirically determine if there were administration mode differences between the traditional paper-and-pencil administration and an online administration of some of the EOC tests, one of which was Algebra I. There was no requirement to participate either in online testing or in the comparability study. Schools could choose either mode of administration, and schools and districts volunteered to participate in the comparability study. Students who participated in the comparability study took their respective EOC tests in both modes.

Because motivation is often cited as a factor in studying student performance, particularly at the high school level where students are more test-wise, one of the benefits to students and schools was that for the purposes of statewide accountability, students would keep whichever score was higher. Students were not notified of their score on their first administration. In order to minimize other confounds, certain restrictions were placed on the study participants. To minimize the effect of additional instruction, students in the comparability study were retested in the different mode no more than 7 school days after the first test administration. Since for the most part, the Algebra I test was administered during the "final exam" period, generally the students had completed instruction in all of their coursework. In order to minimize the effect of familiarity with the test (such as the manner in which topics are addressed or the way questions may be worded), the modes of administration were counterbalanced so that approximately equal numbers of students would be tested in the online environment first, and half would be tested in the traditional paper-and-pencil format first. For the most part, schools adhered to their assigned order (nearly 80%).

Six comparability hypotheses were examined. Results and conclusions follow the statement of each hypothesis.

1. Test content must be the same.
   - As the test content from the paper-and-pencil format was transferred directly to the online format, the test content was the same between modes.
2. The two test modes should have the same measurement precision (i.e., reliability).
   - The reliability coefficients for the paper-and-pencil format ranged from 0.91 to 0.93; for the computer-based administration, the range was 0.92 to 0.93. Thus, the two modes exhibited the same reliability.
3. The two tests should have the same predictive validity coefficients.
   - The relationships between student scores on three other tests and the Algebra I test were examined: grade 8 Math, grade 8 Reading, and Computer Skills. The Algebra I and Math scores correlated 0.84 in the paper-and-pencil mode and 0.85 in the computer-based mode. The Algebra I and Reading scores correlated 0.68 in the paper-and-pencil mode and 0.69 in the computer-based mode. Finally, the Algebra I and Computer Skills scores correlated 0.68 in the paper-and-pencil mode and 0.69 in computer-based mode. The conclusion is that the two modes have the same relationship to external variables used for determining predictive validity.
4. The intercorrelation between the two test scores, corrected for unreliability, should be unity.
   - Fewer than 100 students per form took the same form in both paper-and-pencil and computer-based delivery mode. The correlations between the two sets of raw scores ranged from 0.88 to 0.92. Once these test-retest reliabilities are corrected for the reliability of the test, the intercorrelations range from 0.96 to 1.00.
5. Test scores should yield the same average scale score (within a linear transformation).
   - The mean and standard deviation of the scale score for the paper-and-pencil administration was 154.62 (10.42); for the computer-based administration these moments were 153.54 (10.59). Although the resulting *t*-test gives a statistically significant result, the effect size (*d*) was only -0.10. Generally, an effect size of | 0.20 | is considered to be a small effect.
6. The two tests should have the same factor structure.
   - Confirmatory factor analysis supports the conclusion that the tests have the same factor structure in both delivery modes. Additionally, the tested models for parallelism and tau-equivalence have significant goodness-of-fit chi-squares

($p < 0.05$), and the standardized root mean square residuals for model fit are less than 0.15.

Although there was a slight difference in the mean scale score, the preponderance of evidence suggests that the two modes of test delivery are in fact comparable. Research on this very important topic is continuing for the duration of an Enhanced Assessment Grant from the U.S. Department of Education.

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 EOG and EOC Tests of Mathematics, 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 constructs or concepts measured by the North Carolina EOG Tests of Mathematics are categorized by five basic strands: number and operations, measurement, geometry, data analysis and probability, and algebra. Each item developed for the North Carolina EOG Tests of Mathematics is written to measure one of those five constructs.

Algebra I, Algebra II, and Geometry comprise the EOC Tests of Mathematics. These tests measure the different levels of mathematics knowledge, skills, and abilities specific to the three areas with particular focus on assessing students' ability to process information and engage in higher-order thinking. These elements of mathematics measured by the North Carolina EOC Tests are also categorized into strands: number and operations, measurement and geometry, data analysis and probability, and algebra.

For test-specification summaries, see Appendix B.

Almost all of the items are written by North Carolina teachers and other educators. Some of the math items were written under a contract with a major testing company to handle the logistics, but that contract specified that at least half of the items be written by teachers from North Carolina. Additionally, the items were all reviewed by North Carolina teachers. The contractor-supplied items were actually written for the second edition of the mathematics tests; items that were unused in the previous curriculum cycle were reviewed by content and curriculum experts for a match to the new curriculum. Where possible, the old items were recategorized into the appropriate grade or subject, goal, and objective. Additional items, representing the vast majority of the items written for the third edition, were written by North Carolina educators.

All item writers undergo a day-long training during which time they are taught certain rules and guidelines for item (stem and foil) construction. The training also includes information on Universal Design and access issues for special populations such as students with disabilities and English language learners. The Universal Design training also includes a discussion of considerations for how various accommodations could impact the validity of an item and how to construct items to avoid invalidating an item. Finally, the item writer training includes information on what constitutes bias and how to minimize differential item functioning through careful item construction.

Additionally, all items written are reviewed by at least two content-area teachers from North Carolina. Because North Carolina educators not only deliver the *Standard Course of Study* every day in their classrooms, they 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.

During the review process, the items are also reviewed by a specialist in Exceptional Children and a specialist in English as a Second Language. The specialist teachers review the items in a team with the content teachers in order to make the items as accessible as possible to all populations while preserving the integrity of the curriculum.

The state's teachers are also involved in other aspects of item development and test review (refer to Figure 1: the test development process).

Instructional Validity

DPI routinely administers questionnaires to teachers in an effort to evaluate the validity and appropriateness of the North Carolina End-of-Grade and End-of-Course Tests of Mathematics. 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 (Subject / Grade X) Mathematics curriculum as outlined on the enclosed list of (Subject / Grade X) Mathematics objectives.
2. The test content reflects the goals and objectives of the (Subject / Grade X) Mathematics curriculum as (Subject / Grade X) 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 34: Instructional validity of the content of the North Carolina Tests of Mathematics

Statement	Grade or Subject	% indicating to a superior or high degree
1 Test alignment to *SCS*	Pretest—Grade 3	100
	3	83
	4	100
	5	100
	6	100
	7	100
	8	80
	Algebra II	88
2 Test alignment to instruction	Pretest—Grade 3	93
	3	63
	4	74
	5	76
	6	92
	7	100
	8	89
	Algebra II	81
3 Item clarity and vocabulary	Pretest—Grade 3	79
	3	79
	4	74
	5	82
	6	100
	7	100
	8	89
	Algebra II	94
4 Content and demographic balance	Pretest—Grade 3	93
	3	100
	4	81
	5	82
	6	100
	7	100
	8	100
	Algebra II	88
5 Distractor design and suitability	Pretest—Grade 3	86
	3	89
	4	89
	5	94
	6	100
	7	100
	8	83
	Algebra II	81

The Algebra I and Geometry *Standard Courses of Study* had more changes from the second to third edition than did Algebra II. In spite of extensive professional development provided by the NCDPI Instructional Services division and allowing an extra transition year before full implementation, some of the changes still were unexpected. In addition to responding to the five instructional validity statements noted, teachers also had the opportunity to provide comments. In Algebra I, a substantial amount of matrix algebra was moved from Algebra II and there was an increased emphasis on linear models. For Geometry, teachers appeared to have expected a different balance of goal 2 and goal 3 items than what was in the test specifications and did not expect the degree of integration of Algebraic concepts into Geometric topics, including the use of matrices. In both test reviews there were 13 teachers, so a response from 1 teacher is approximately 9%. Not all teachers responded to all statements.

Table 35: Instructional validity of the content of the North Carolina EOC Tests of Algebra I and Geometry

Statement	**Course**	**% indicating to a superior or high degree**	**% indicating to an average degree**	**% indicating to a low degree**
1 Test alignment to SCS	Algebra I Geometry	73 33	27 42	0 25
2 Test alignment to instruction	Algebra I Geometry	50 15	50 54	0 31
3 Item clarity and vocabulary	Algebra I Geometry	54 92	46 8	0 0
4 Content and demographic balance	Algebra I Geometry	92 91	0 9	8 0
5 Distractor design and suitability	Algebra I Geometry	82 83	9 17	9 0

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 EOG and EOC Tests of Mathematics, 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 EOG and EOC Tests of Mathematics range from 0.47 to 0.81 indicating a moderate to strong correlation between EOG and EOC scale scores and variables based on teacher judgment of the students' attainment of the content.* The tables below provide the Pearson correlation coefficients for variables used to establish criterion-related validity for the North Carolina EOG and EOC Tests of Mathematics.

**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 36: Pearson correlation coefficient table for variables used to establish criterion-related validity for the North Carolina EOG Tests of Mathematics

Grade	Pre 3	3	4	5	6	7	8
Teacher Judgment of Achievement Level by Assigned Achievement Level	0.49	0.63	0.62	0.62	0.62	0.62	0.64
Teacher Judgment of Achievement by Expected Grade	0.71*	0.78	0.77	0.73	0.69	0.66	0.66
Teacher Judgment of Achievement by Math Scale Score	0.52	0.66	0.65	0.65	0.65	0.65	0.66
Assigned Achievement Level by Expected Grade	0.47*	0.60	0.58	0.57	0.56	0.52	0.52
Expected Grade by Math Scale Score	0.50*	0.63	0.61	0.61	0.59	0.56	0.55

*The correlate is the actual final grade the student earned in the second grade.

As might be expected, the correlations for the Pretest—Grade 3 were lower than the corresponding relationships at the higher grades. Although teachers were asked to consult with the student's second-grade teacher to qualify the student's achievement, in many cases this was not possible. Anecdotal evidence suggests that some teachers attempted to answer these questions based on their own knowledge of the student for the first week or two of school.

Table 37: Pearson correlation coefficient table for variables used to establish criterion-related validity for the North Carolina EOC Tests of Mathematics

	Algebra I	Geometry	Algebra II
Teacher Judgment of Achievement Level by Assigned Achievement Level	0.63	0.64	0.59
Teacher Judgment of Achievement by Expected Grade	0.79	0.82	0.79
Teacher Judgment of Achievement by Scale Score	0.65	0.67	0.63
Assigned Achievement Level by Expected Grade	0.60	0.62	0.57
Expected Grade by Scale Score	0.62	0.64	0.60

The variables used in the tables above are as follows:

- **Teacher Judgment of Achievement Level**: 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 3-4.
- **Assigned Achievement Level**: The achievement level assigned to a student based on his or her test score, based on the cut scores previously described on page 45.
- **Expected Grade**: Teachers were also asked to provide for each student the letter grade that they anticipated each student would receive at the end of the grade or course. As noted, the Pretest—Grade 3 variable was actually the student's final grade 2 grade.
- **Scale Score**: The converted raw-score-to-scale-score value obtained by each examinee.

DPI found moderate to strong correlations between scale scores in mathematics and variables such as teachers' judgment of student achievement, expected grade, and assigned achievement levels (all measures of concurrent validity). Equally important is to demonstrate that the test scores are ***not*** correlated with external variables that should not be related to student proficiency or should only be moderately correlated. The department also found generally low correlations among the test scores and variables external to the test such as gender, limited English proficiency, and disability for all grades and subjects.

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

Table 38: Tetrachoric correlation coefficient table for additional presumably uncorrelated variables used to establish criterion-related validity for the North Carolina Tests of Mathematics

Score by Grade/Subject	× Gender	× LEP	× Disability	× Migrant	× FRL	× Title I
Pretest—Grade 3	0.00	-0.32	-0.22	-0.12	-0.33	-0.29
Grade 3	0.03	-0.25	-0.33	-0.09	-0.44	-0.29
Grade 4	0.02	-0.23	-0.36	-0.14	-0.43	-0.26
Grade 5	0.02	-0.24	-0.41	-0.17	-0.43	-0.26
Grade 6	-0.01	-0.28	-0.42	-0.20	-0.46	-0.19
Grade 7	-0.02	-0.27	-0.44	-0.18	-0.45	-0.19
Grade 8	-0.03	-0.24	-0.44	-0.13	-0.45	-0.19
Algebra I	-0.02	-0.24	-0.41	-0.17	-0.38	0.07
Geometry	0.06	-0.18	-0.24	-0.15	-0.35	-0.15
Algebra II	-0.02	-0.24	-0.41	-0.17	-0.38	0.07

Nearly half (45%) of the correlations between scores and external variables were less extreme than ± 0.2, and only 20% of the correlations between scores and external variables were more extreme than ± 0.4. None of these relationships exceeded the degree of relationship recorded for the selected measures of concurrent validity. These generalizations held across the full range of forms administered by DPI for all the grades and subject areas.

7.4 Concurrent and Predictive Validity

Concurrent validity and predictive validity can be demonstrated if a test's scores or interpretations show a strong relationship to the scores or interpretations from another, already validated instrument that measures the same construct or a closely related construct. Conclusions about concurrent validity information, as the name suggests, can be drawn when the two measures occur at or nearly at the same time. Predictive validity, on the other hand, would imply the earlier measurement provides information on the performance of the test-taker on the second measure at a later point in time.

Because the North Carolina tests are the only tests that measure the North Carolina *Standard Courses of Study*, it is difficult to obtain obvious concurrent validity data. Instead, concurrent validity in this situation must be inferred from other tests of general mathematical reasoning and problem solving.

One portion of the SAT Reasoning Test measures mathematical problem-solving, analysis, and critical-thinking skills. Thus, if there is a strong relationship between SAT Mathematics scores and EOG or EOC mathematics scale scores, some degree of validity can be imputed. Although the state education agencies do not receive individual student scores back from the SAT, local education agencies do. Taken in aggregate, how well an LEA performs on the EOG or EOC tests can be correlated to how well the same LEA performs on the SAT. One drawback to looking at this aggregated data is that it does not necessarily compare the same student's performance on the EOG or EOC test to his or her performance on the SAT. Although North Carolina has a high proportion of high school students taking the SAT, even those who are not planning on entering a four-year college or university, there is still a degree of self-selection among the students who opt to participate in the SAT. This restriction in range of examinee ability can also mask evidence of a relationship between the two variables. Additionally, varying participation rates among LEAs can also dilute the relationship. Even with these limitations, the relationships between LEA EOG or EOC scores and the LEA's SAT scores are quite high.

Table 39: Correlations between SAT scores and EOG or EOC scores, aggregated at the LEA level

	Grade 8 EOG Math	Algebra I	Geometry
SAT Math	0.82	0.77	0.83
SAT Critical Reading	0.78	0.71	0.75

These coefficients can be viewed as predictive validity evidence; because many students take Geometry in the 10th or 11th grade, the relationship between Geometry scores and SAT scores may also have a component of concurrent validity. Discriminant validity is shown by the weaker correlations with the SAT Critical Reading score in each comparison.

Another source of concurrent validity is the trend between students' progress on the National Assessment of Education Progress (NAEP) and their progress on end-of-grade scores. Although the scores themselves cannot and should not be compared directly, nor is it valid to compare the percent "proficient" on each test, the trends show corresponding increases in both NAEP math scores and scores on the North Carolina EOG Tests of Mathematics in previous editions.

Figures 38 through 41 show the trends for students who score "basic" or "proficient" on NAEP assessments in grades 4 and 8 compared to students who scored at Level III or above on the North Carolina End-of-Grade Tests of Mathematics in grade 4 and 8. The NC EOG data points up to 2006, although not connected in the graphs below, may be viewed as establishing a trend. Note that the NC EOG data points are not connected; this is to emphasize that in 2006, new, more rigorous math achievement level standards (cut scores) were adopted with the new test edition, thereby breaking the trend. It is fully expected that a similar pattern of growth will be evident in future years.

Figure 38: Comparison of NAEP "proficient" scores and North Carolina End-of-Grade Tests of Mathematics scores for Grade 4

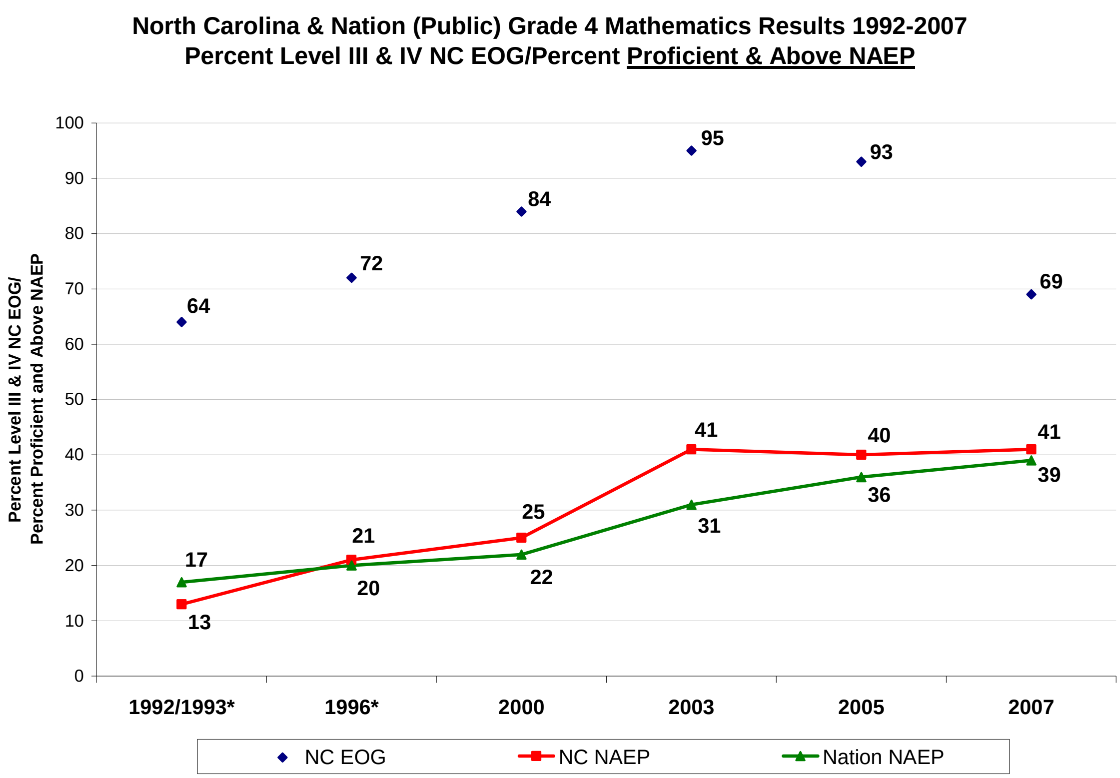


Figure 39: Comparison of NAEP "basic" scores and North Carolina End-of-Grade Tests of Mathematics scores for Grade 4

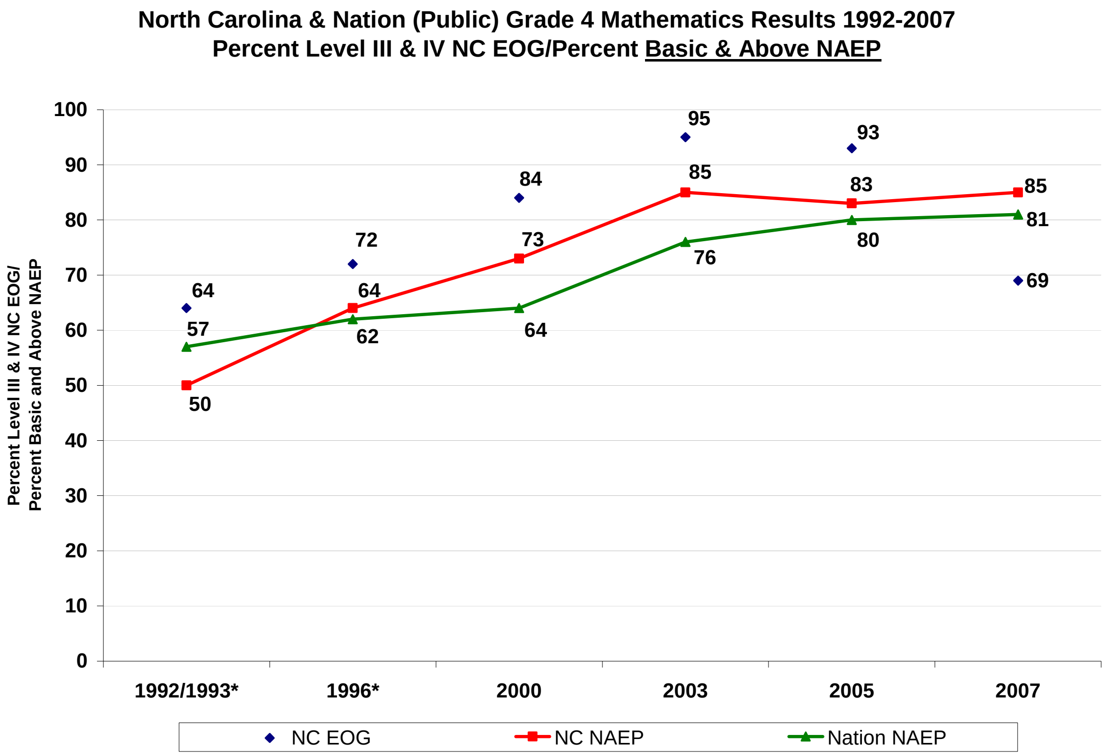


Figure 40: Comparison of NAEP "proficient" scores and North Carolina End-of-Grade Tests of Mathematics scores for Grade 8

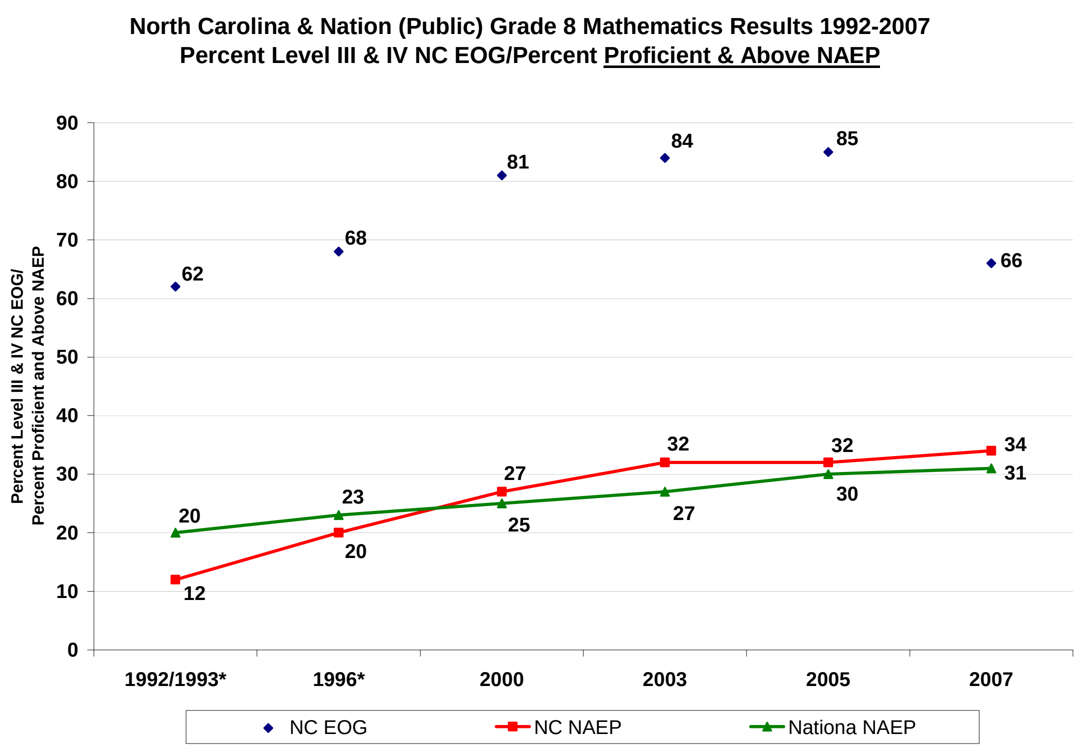
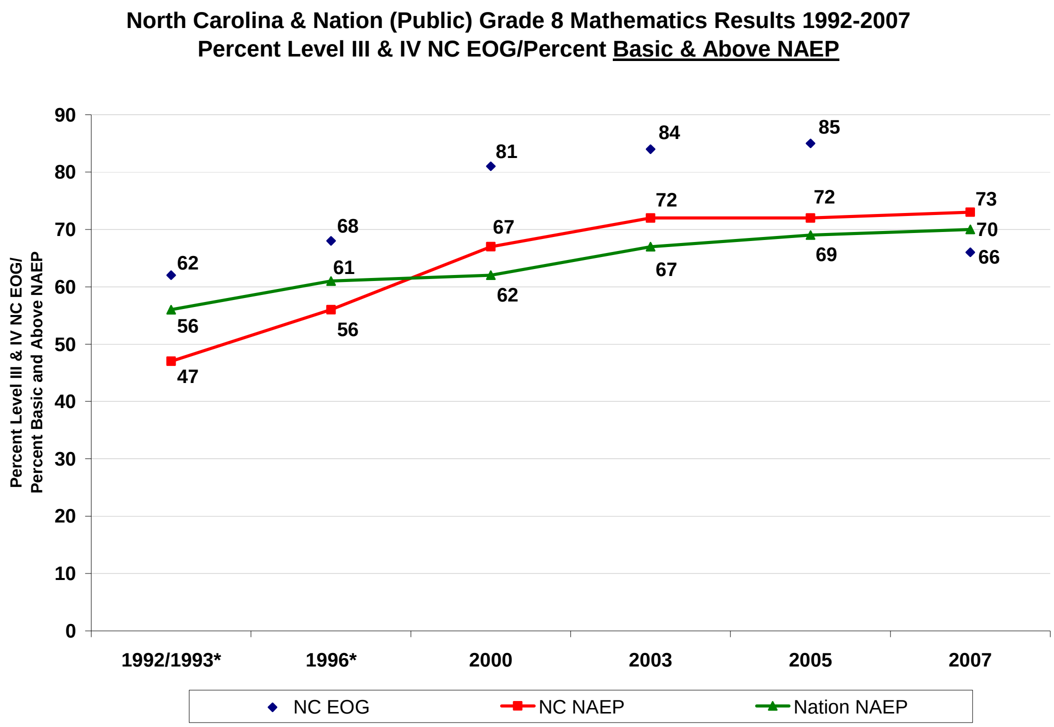


Figure 41: Comparison of NAEP "basic" scores and North Caroline End-of-Grade Tests of Mathematics scores for Grade 8



7.5 Alignment

A final element to ensure test validity is that the test is closely aligned to the content it is intended to measure. The NCDPI contracted with an outside provider to conduct alignment studies of the edition 3 NC EOG Tests of Mathematics to the 2003 Mathematics *Standard Course of Study*. The report is not yet complete as of this writing, so results presented here are from a draft provided to the NCDPI by the contractor. A more complete treatment of the alignment study results is provided in Appendix G.

Four elements of alignment were quantified by the panelists:

Categorical Concurrence: *The categorical-concurrence criterion provides a very general indication of alignment if both documents incorporate the same content.* The criterion of categorical concurrence between goals and assessment is met if the same or consistent categories of content appear in both documents. *This criterion was judged by determining whether the assessment included items measuring content from each goal. The analysis assumed that the assessment had to have at least six items measuring content from a goal in order for an acceptable level of categorical concurrence to exist between the goal and the assessment.*

Depth of Knowledge: *Depth-of-knowledge consistency between goals and assessment indicates alignment if what is elicited from students on the assessment is as demanding cognitively as what students are expected to know and do as stated in the goals.* For consistency to exist between the assessment and the goal, as judged in this analysis, at least 50% of the items corresponding to a goal had to be at or above the level of knowledge of the goal; 50%, a conservative cutoff point, is based on the assumption that a minimal passing score for any one goal of 50% or higher would require the student to successfully answer at least some items at or above the depth-of-knowledge level of the corresponding goal.

Range of Knowledge: The range-of-knowledge criterion is used to judge whether a comparable span of knowledge expected of students by a goal is the same as, or corresponds to, the span of knowledge that students need in order to correctly answer the assessment items/activities. *The criterion for correspondence between span of knowledge for a goal and an assessment considers the number of standards within the goal with one related assessment item/activity. Fifty percent of the standards for a goal had to have at least one related assessment item in order for the alignment on this criterion to be judged acceptable.*

Balance of Representation: *The balance-of-representation criterion is used to indicate the degree to which one standard is given more emphasis on the assessment than another.* An index is used to judge the distribution of assessment items. The index is computed by considering the difference in the proportion of standards and the proportion of hits assigned to the standard.

In general, the results of the alignment study showed strong alignment on the range-of-knowledge criterion. One form at grade 5 was determined to have a "weak" balance of representation, but otherwise all forms at all grades showed strong alignment on this criterion. For the criterion of categorical concurrence, strong alignment was found in all but three circumstances. At grades 5 and 7, the panelists noted that there were fewer than 6 items for goal 2, measurement. However, the tests do conform to the recommendations of the test specifications committee, in which 10% to 15% of the items were to come from this goal. Additionally, the alignment study panelists felt that one of the grade 5 forms did not have sufficient items for goal 4, data analysis and probability.

Because of the methodology used to examine alignment to the depth-of-knowledge criterion, it appears that the NC EOG Tests of Mathematics either have weak alignment or do not meet the criterion as described. It is important to note, however, that the criterion used in the study was that at least half of the items on the test forms had to be at or *above* the depth-of-knowledge level called for in the *Standard Course of Study*. The NCDPI feels that this criterion is poorly articulated in the context of a high-stakes educational achievement test. Each goal and objective in the *Standard Course of Study* can be mapped to a particular level of thinking, as described in Chapter 2, and, as in the alignment study, to a depth-of-knowledge level. The practice of the NCDPI is to not to exceed the expectation outlined in the *SCS*. Thus the NCDPI intentionally does *not* ask questions on the EOG and EOC tests that exceed the indicated thinking skill. The rationale is that the *SCS* states the knowledge, skills, and abilities that must be taught, therefore, to then ask the students to perform tasks on the EOG or EOC tests that are beyond the expectations of the *SCS* is inappropriate. However, based on the results of the alignment study, the NCDPI has instituted an additional item-development criterion: that the bulk of the items are written at the maximum thinking skill for each objective, with a goal of at least half the items on the test being at the maximum thinking skill.

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. After the review panel has approved a test form, test forms are then configured to go through the printing process. A PDF file is sent directly to the printer when the final approval of the camera-ready copy has been obtained. Once all test answer sheets and booklets are printed, the operations specialist from the NCDPI and the warehouse manager conduct 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 pre-code 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/NC WISE (Student Information Management System / North Carolina Window of Information for Student Education) database. Precoded answer sheets provide schools with the opportunity to correct or update information in the SIMS/NC WISE 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 the 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 the sheets 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 an asterisk (*) indicating that the student has answered twice.

8.4 Quality Control of Test Scores and Data Merging

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 is in, the NCDPI runs a CHECK KEYS program to flag areas where answer keys may need a second check.

As data come into the NCDPI, Student Information and Accountability Systems (a division of Technology and Information Services) staff import and clean data to ensure that individual student files are complete. Additionally, certain student demographic information is merged into the test data files from authoritative sources. For example, student Free and Reduced Lunch status is imported from the School Nutrition data collection activity. Other demographic variables that are imported from other data collections throughout the year are gender, ethnicity, and LEP status.

8.5 Quality Control in Reporting

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

"The department hereby certifies the accuracy of the data from the North Carolina end-of-course tests for Fall 2007 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 2007 semester."

Glossary of Key Terms

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

Accommodations		Changes made in the format or administration of the test to provide options to test takers who are unable to take the original test under standard test conditions.
Achievement levels		Descriptions of a test taker's competency in a particular area of knowledge or skill, usually defined as ordered categories on a continuum classified by broad ranges of performance.
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. (For math it is generally 0.15 and for social studies it is generally 0.22).
Biserial correlation		The relationship between an item score (right or wrong) and a total test score.
Cut scores		A specific point on a score scale, such that scores at or above that point are interpreted or acted upon differently from scores below that point.
Dimensionality		The extent to which a test item measures more than one ability.
Embedded test model		Using an operational test to field test new items or sections. The new items or sections are "embedded" into the new test and appear to examinees as being indistinguishable from the operational test.
Equivalent forms		Statistically insignificant differences between forms (i.e., the red form is not harder).
Field test		A collection of items to approximate how a test form will work. Statistics produced will be used in interpreting item behavior/performance and allow for the calibration of item parameters used in equating tests.

Foil counts		Number of examinees that endorse each foil (e.g., number who answer "A," number who answer "B," etc.)
Item response theory		A method of test item analysis that takes into account the ability of the examinee and determines characteristics of the item relative to other items in the test. The NCDPI uses the 3-parameter model, which provides slope, threshold, and asymptote.
Item tryout		A collection of a limited number of items of a new type, a new format, or a new curriculum. Only a few forms are assembled to determine the performance of new items and not all objectives are tested.
Mantel-Haenszel		A statistical procedure that examines the differential item functioning (DIF) or the relationship between a score on an item and the different groups answering the item (e.g., gender, race). This procedure is used to identify individual items for further bias review.
Operational test		Test is administered statewide with uniform procedures and 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		Test forms cover 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.
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 test and arrangement of objectives
Threshold		The point on the ability scale where the probability of a correct response is 50%. Threshold for an item of average difficulty is 0.00.

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.

Gregory, Robert J. (2000). *Psychological Testing: History, Principles, and Applications.* Needham Heights: Allyn & Bacon.

Hambleton, Ronald K. (1983). *Applications of Item Response Theory*. British Columbia: Educational Research Institute of British Columbia.

Hinkle, D.E., Wiersma, W., & Jurs, S. G. (1998). *Applied Statistics for the Behavioral Sciences* (pp. 69–70)

Kolen, M.J., & Brennan, R.L. (1995). *Test equating methods and practices.* New York: Springer.

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.

Mitzel, H. C., Lewis, D. M., Patz, R. J., & Green, D. R. (2001). The bookmark procedure: Psychological perspectives. In G. J. Cizek (Ed.), *Setting performance standards:* Concepts, methods, and perspectives (pp. 249–281). Mahwah, NJ: Erlbaum.

Muraki, E., Mislevy, R.J., & Bock, R.D. (1991). PC-BiMain: Analayis of item parameter drift, differential item functioning, and variant item performance [Computer software]. Mooresville, IN: Scientific Software, Inc.

Roeber, E. (2002). *Setting standards on alternate assessments* (Synthesis Report 42). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Retrieved [December 13, 2007], from the World Wide Web: http://education.umn.edu/NCEO/OnlinePubs/Synthesis42.html

Thissen, D., & Orlando, M. (2001). Item response theory for items scored in two categories. In D. Thissen & H. Wainer (Eds), *Test Scoring* (pp. 73–140). Mahwah, NJ: Lawrence Erlbaum Associates.

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.

Additional Resources

Anastasi, A. (1982). *Psychological Testing*. New York: Macmillan Publishing Company, Inc.

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.

Camilli, G. & Shepard, L.A. (1994). *Methods for Identifying Biased Test Items*. Thousand Oaks, CA: Sage Publications, Inc.

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.

Dorans, N.J. & Holland, P.W. (1993). DIF Detection and description: Mantel-Haenszel and standardization. In P.W. Holland and H. Wainer (Eds.), *Differential Item Functioning* (pp. 35–66). Hillsdale, NJ: Lawrence Erlbaum.

Haladyna, T.M. (1994). *Developing and Validating Multiple-Choice Test Items*. Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Hambleton, R.K. & Swaminathan, H. (1985). *Item Response Theory: Principles and Applications.* Kluwer-Nijhoff Publishing.

Hambleton, R.K., Swaminathan, H., & Rogers, H.J. (1991). *Fundamentals of Item Response Theory*. Newbury Park, CA: Sage Publications, Inc.

Holland, P.W. & Wainer, H. (1993). *Differential Item Functioning*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Joreskog, K.J. & Sorbom, D. (1986). PRELIS: A program for multivariate data screening and data summarization. Chicago, IL: Scientific Software, Inc.

Joreskog, K.J. & Sorbom, D. (1988). LISREL 7: A guide to the program and applications. Chicago, IL: SPSS, Inc.

Kubiszyn, T. & Borich, G. (1990). *Educational Testing and Measurement*. New York: HarperCollins Publishers.

Muraki, E., Mislevy, R.J., & Bock, R.D. PC-Bimain Manual. (1991). Chicago, IL: Scientific Software, Inc.

National Council of Teachers of Mathematics. Curriculum and Evaluation Standards for School Mathematics. (1989). Reston, VA: Author.

North Carolina Department of Public Instruction. (1992). Teacher Handbook for Mathematics. Raleigh, NC: Author.

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.

SAS Institute, Inc. (1985). The FREQ Procedure. In SAS User's Guide: Statistics, Version 5 Edition. Cary, NC: Author.

Traub, R.E. (1994). *Reliability for the social sciences: Theory and applications*. Thousand Oaks, CA: Sage Publications, Inc.

Appendix A – Item-Development Guidelines

Procedural Guidelines

1. Use the best-answer format.
2. Avoid writing complex multiple-choice items.
3. Format the items vertically, not horizontally.
4. Avoid errors of grammar, abbreviation, punctuation, and spelling.
5. Minimize student reading time.
6. Avoid tricky or misleading items.
7. Avoid the use of contractions.
8. Avoid the use of first or second person.

Content Guidelines

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

Stem Construction Guidelines

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

General Foil Development

21. Each item must contain four foils (A, B, C, 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 so that phrases in the stem (clang associations) are repeated in the foils.
    Also avoid including ridiculous options.
28. Avoid grammatical clues to the correct answer.
29. 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 or italic.

30. The correct response for items written should be evenly balanced among the response options. For a four-option multiple-choice item, correct responses should be located at each option position about 25 percent of the time.
31. The items written should contain one and only one best (correct) answer.

Distractor Development

32. 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.
33. To the extent possible, use the common errors made by students as distractors.
34. Technically written phrases may be used, where appropriate, as plausible distractors.
35. True phrases that do not correctly respond to the stem may be used as plausible distractors where appropriate.
36. The use of humor should be avoided.

Appendix B – Test-Specification Summaries

Some sections of the test-specifications documents are the same for some or all of the grades and subjects. In presenting the test-specifications documents in this technical manual, repetitive information has been deleted in subsequent grades and the omission noted.

The test specifications were developed and approved in early 2003. Some refinements were made based on the outcome of field testing and subsequent discussions with staff from both Testing / Accountability and Curriculum / Instructional Services. Additionally, some information, such as web links, has become outdated. Substantial corrections are noted below.

The URL for the math indicators has changed since these test-specifications documents were approved. The link in section 6 below is no longer correct. The current URL is http://community.learnnc.org/dpi/math/archives/instructional_resources/ and the indicators themselves can be found at the links to "Grades 3–5 Resources (2003 *SCS*)," "Grades 6–8 Resources (2003 *SCS*)," and "Grades 9–12 Resources (2003 *SCS*)."

After these test-specifications documents were approved, the decision was also made to not have a per-section time limit, as described in section 21. Instead, administration times were calculated to apply to an entire part of the test, calculator active or calculator inactive. Finally, a decision was subsequently made to have the grade 6 and 7 math tests be the same length as the tests in grades 3 through 5.

The North Carolina Pretest of Mathematics—Grade 3
Aligned to the 2003 Mathematics Curriculum
Test Specifications

Design of Assessments

North Carolina assessments are designed to be accessible to as many students as possible. Item writers will be trained on the universal design principles to ensure that items are accessible to Exceptional Children and Limited English Proficient students. The NCDPI consultants representing the needs of Exceptional Children and Limited English Proficient students will be expected to work with NCDPI Testing and Curriculum staff in reviewing and editing test items. Special consideration will be given to how items will be read aloud to students requiring a "read-aloud" accommodation and for how items will be signed to deaf or hearing-impaired students.

Items must be accessible to students taking the test using regular-print test booklets, large-print test booklets, Braille editions, or a computer screen. **Items that cannot be adapted to fit into any one or more of these modes will not be used on the assessment.**

Purpose and Uses

The North Carolina Pretest of Mathematics—Grade 3 will be administered to students at the beginning (within the first three weeks of school) of grade 3. The grade 3 pretest discussed in this document will measure the grade 2 goals and objectives found in the 2003 North Carolina Mathematics *Standard Course of Study*. The pretest will provide prescores for students at the beginning of grade 3 for the ABCs accountability program. Grade 3 prescores are necessary to provide predata for the growth analysis for students at the end of grade 3.

Timeline for Aligning Tests to 2003 Mathematics Curriculum

Year	Activity
2003–2004	Test Specifications, Item Development
2004–2005	Embedding Study
2005–2006	Field Test (paper/pencil and online)
Summer/fall 2006	Operational

2003 Mathematics *Standard Course of Study* and Grade 2-Level Competencies

Major Concepts/Skills	Concepts/Skills to Maintain
• Number sense 0–999	• Patterns
• Place value	• Sort and classify
• Addition and subtraction of multidigit numbers	• Line plots, tallies
• Length, time	
• Symmetry and congruence	
• Pictographs	
• Probability experiments	
• Number sentences	
• Students will solve relevant and authentic problems using appropriate technology and apply these concepts as well as those developed in earlier years	

Strands: Number and Operations, Measurement, Geometry, Data Analysis and Probability, Algebra

COMPETENCY GOAL 1: The learner will read, write, model, and compute with whole numbers through 999.

Objectives

1.01 Develop number sense for whole numbers through 999.

a. Connect model, number word, and number using a variety of representations.
b. Read and write numbers.
c. Compare and order.
d. Rename.
e. Estimate.
f. Use a variety of models to build understanding of place value (ones, tens, hundreds).

1.02 Use area or region models and set models of fractions to explore part-whole relationships in contexts.

a. Represent fractions (halves, thirds, fourths) concretely and symbolically.

b. Compare fractions (halves, thirds, fourths) using models.
c. Make different representations of the same fraction.
d. Combine fractions to describe parts of a whole.

1.03 Create, model, and solve problems that involve addition, subtraction, equal grouping, and division into halves, thirds, and fourths (record in fraction form).

1.04 Develop fluency with multidigit addition and subtraction through 999 using multiple strategies.
a. Strategies for adding and subtracting numbers.
b. Estimation of sums and differences in appropriate situations.
c. Relationships between operations.

1.05 Create and solve problems using strategies such as modeling, composing and decomposing quantities, using doubles, and making tens and hundreds.

1.06 Define and recognize odd and even numbers.

COMPETENCY GOAL 2: The learner will recognize and use standard units of metric and customary measurement.

Objectives

2.01 Estimate and measure using appropriate units.
a. Length (meters, centimeters, feet, inches, yards).
b. Temperature (Fahrenheit).

2.02 Tell time at the five-minute intervals.

COMPETENCY GOAL 3: The learner will perform simple transformations.

Objectives

3.01 Combine simple figures to create a given shape.

3.02 Describe the change in attributes as two- and three-dimensional figures are cut and rearranged.

3.03 Identify and make:
a. Symmetric figures.
b. Congruent figures.

COMPETENCY GOAL 4: The learner will understand and use data and simple probability concepts.

Objectives

4.01 Collect, organize, describe and display data using Venn diagrams (three sets) and pictographs where symbols represent multiple units (2's, 5's, 10's).

4.02 Conduct simple probability experiments; describe the results and make predictions.

COMPETENCY GOAL 5: The learner will recognize and represent patterns and simple mathematical relationships.

Objectives

5.01 Identify, describe, translate, and extend repeating and growing patterns.

5.02 Write addition and subtraction number sentences to represent a problem; use symbols to represent unknown quantities.

Test format

The test will consist of two parts as shown in the table below.

Calculator Inactive Part	Calculator Active Part
No calculator allowed	Calculator allowed
30% of test	70% of test
Assess Strand 1 only	Assess all Strands
Ruler, graph paper, scratch paper allowed	Ruler, graph paper, scratch paper allowed

Objectives Not Tested

Strand ("Goal")	Objective(s)	Remarks
1: Number and Operations	1.01e (number sense through estimation), 1.04a (strategies for addition, subtraction), 1.04c (relationship between addition and subtraction)	Estimation of sums, differences tested as 1.04b; addition and subtraction tested (see especially 1.03), but not specific algorithms ("carrying", "borrowing"); addition and subtraction of 3-digit numbers tested in calculator active section ***only***
2: Measurement*	All objectives tested.	
3: Geometry*	3.01 (combine simple figures), 3.02 (changes in attributes after cutting and rearranging)	some solid geometry ("3-D") concepts and vocabulary from grade 1 may be used to provide context in items assessing other grade 2 objectives
4: Data Analysis and Probability	All objectives tested.	
5: Algebra	5.02 (write number sentences to represent a problem; use symbols to represent unknown quantities)	

Clarifications of Objectives Needed for Item Development

Objective	Remarks
1.01d	focus on grouping into groups of tens (573 contains 57 groups of ten)
1.02	fractions: models must be used; emphasize discrete rather than continuous models; do not use improper fractions
1.03	use term "sharing"; avoid referring to "division" or "dividing"
1.04b	estimating sums, differences: use ranges as foils
1.06	odd and even numbers: focus on recognition, not definition
2.01a	explicitly state when ruler is to be used; measure to the whole unit; students should be expected to turn (rotate) the ruler while measuring
2.01b	thermometers: both 1- and 2-degree intervals should be used
3.03	knowledge of terms "symmetric" and "congruent" is expected
4.01	Venn diagrams: up to three sets can be used; include data from science or social studies
4.02	specify "fair" spinner; may also use "number cubes"; avoid coins and disks; include data from science or social studies
5.01	"growth" patterns involving subtraction are acceptable

Level of Difficulty
Prior to field testing, North Carolina educators will use professional judgments to place an item into one of three categories: Easy, Medium, or Hard. After field testing, the items will be placed into one of the same three categories based on statistics related to student performance.

Level of Difficulty	Percent of Items per Form
Easy	25%
Medium	50%
Hard	25%

Thinking Skills
The thinking skills framework used in developing the grade 3 pretest is from *Dimensions of Thinking* by Robert J. Marzano and others (1988). Items are categorized as requiring lower-order or higher-order thinking skills.

Lower-Order Thinking Skills	Higher-Order Thinking Skills
Knowledge	Analyzing
Organizing	Generating
Applying	Integrating
	Evaluating
40% per form	**60% per form**

Item Formats
All items will be in multiple-choice format with one and only one correct answer choice and three plausible distractors.

Number of Items per Form
The exact number of items will not be determined until after the embedding study. The test will be designed so that the total number of questions will not exceed 80 items. For example, the test may have 60 operational items and 20 field test items.

Calculator Use
Calculators are only allowed during the Calculator Active part of the test.

Minimum Calculator Requirement
Calculator Inactive: Calculator use is not allowed.
Calculator Active: Four-function calculator with memory key

Calculator Requirements
The use of keyboarding calculators and symbol-manipulation calculators is prohibited during the administration of any North Carolina test. Symbol-manipulation calculators are those capable of doing symbolic algebra (i.e, factoring, expanding, or simplifying given variable output), or symbolic calculus. As curricula and technology change, the policy concerning technology use on North Carolina tests (both multiple-choice and open-ended) will be reevaluated.

Additional Assessment Materials
Students will be allowed to use a ruler, graph paper, and scratch paper. A calculator is allowed on the Calculator Active part of the test only. No formula sheets will be required. No other manipulatives will be required.

Percent of Items by Goal per Test Form
The test will be designed to meet a percent-by-strand specification rather than a number-per-objective specification.

Goal	Percent	Priority of Objective Coverage
Number & Operation	55–60%	1.05, 1.02, 1.03, 1.01, 1.04, 1.06
Measurement	10–12%	Equal priority
Geometry	5–6 %	Equal priority
Data Analysis & Probability	10–14%	4.01, 4.02
Algebra	10–14%	Not needed (only 1 objective tested)

Reporting by Strand
After standards have been set in January of 2007, Goal Summary Reports will be made available. The Goal Summary Report will provide information by strand for Number &

Operation, Data Analysis & Probability, and Algebra. The Measurement and Geometry strands will be combined for reporting purposes.

Special Requests
The following special requests made by the Curriculum, EC, or LEP staff during various meetings will be considered:

1. Avoid using “not” or “except” questions.
2. Provide students with rotating two-sided ruler on screen for NCCATS accommodation.
3. In the test booklet, provide written instructions to “Choose the ***best*** answer.”

Guidelines for Item Development and a Style Manual are available for guiding item development. In addition, the NCDPI consultants representing the needs of Exceptional Children and Limited English Proficient students have written documents outlining “Things to Consider” when writing test items. At the start of each testing year, these documents are updated.

Meetings and Attendance
Preliminary meetings were held to discuss how to align tests to the 2003 curriculum. Decisions made at these meetings helped to guide discussions at the Grade 3 Test Specifications meeting held on May 23, 2003.

Preliminary Meetings to Discuss Aligning Tests to the 2003 Math Curriculum
April 11, 2003, room 694, Education Building, 9 a.m.–11 a.m.
Attendance
Sarah McManus, Lead Consultant for Mathematics Assessments, NCDPI Testing
Laura Kramer, Senior Psychometrician, NCDPI Testing
Bill Tucci, Section Chief for Mathematics and Science, NCDPI Instructional Services
Toni Meyer, Mathematics Consultant, K–5, NCDPI Instructional Services
Bill Scott, Mathematics Consultant, 9–12, NCDPI Instructional Services
Linda Patch, Mathematics Consultant, 6–8, NCDPI Instructional Services

April 30, 2003, 5th floor library, Education Building, 11 a.m.–12:30 p.m.
Attendance
Sarah McManus, Lead Consultant for Mathematics Assessments, NCDPI Testing
Kevin Murphy, Operations Manager, NCDPI Testing
Toni Meyer, Mathematics Consultant, K–5, NCDPI Instructional Services
Bill Scott, Mathematics Consultant, 9–12, NCDPI Instructional Services
Linda Patch, Mathematics Consultant, 6–8, NCDPI Instructional Services
Diann Irwin, Section Chief for Behavior Support Services, NCDPI Exceptional Children
Martha Downing, Consultant for the Hearing Impaired, NCDPI, Exceptional Children
Tom Winton, Consultant for the Visual Impaired, NCDPI, Exceptional Children

May 15, 2003, Room 150, Education Building, 2 p.m.–4 p.m.
Attendance
Lou Fabrizio, Director, Division of Accountability Services, NCDPI
Mildred Bazemore, Section Chief for Test Development, NCDPI
Sarah McManus, Lead Consultant for Mathematics Assessments, NCDPI Testing,

Laura Kramer, Senior Psychometrician, NCDPI Testing
June Atkinson, Director, Division of Instructional Services, NCDPI
Bill Tucci, Section Chief for Mathematics and Science, NCDPI Instructional Services
Toni Meyer, Mathematics Consultant, K–5, NCDPI Instructional Services
Bill Scott, Mathematics Consultant, 9–12, NCDPI Instructional Services
Linda Patch, Mathematics Consultant, 6–8, NCDPI Instructional Services

Test Coordinator's Advisory Committee
May 1, 2003, Room 224, Education Building, 10:45 a.m.–11:15 a.m.
Sarah McManus presented information to the committee about proposed new item formats. Feedback was gathered.

Technical Advisory Committee Meeting
May 8, 2003, SBE lounge, 7th floor, Education Building, 1 p.m.–2 p.m.
Sarah McManus presented information to the committee about proposed new item formats. Feedback was gathered.

Grade 3 Pretest Test Specifications Committee Meeting
May 23, 2003, Room 695, Education Building 9:00 a.m.–2 p.m.
Attendance
Sarah McManus, Lead Consultant for Mathematics Assessments, NCDPI Testing
Dr. Laura Kramer, Senior Psychometrician, NCDPI Testing
Bill Tucci, Section Chief for Mathematics and Science, NCDPI Instructional Services
Toni Meyer, Mathematics Consultant, K–5, NCDPI Instructional Services
Bill Scott, Mathematics Consultant, 9–12, NCDPI Instructional Services
Robert Brown, Education Consultant, Technical Outreach for Public Schools
Sallie Abbas, Mathematics Consultant, Technical Outreach for Public Schools
Terry Gunter, Elementary Mathematics Specialist, Durham County Public Schools
Patricia Jordan, Mathematics Instructional Resource Teacher, Poe Elementary School, Wake

As the test specifications for other mathematics tests are developed and as embedding plans continue to be discussed, portions of this document (i.e. number of items) will need to be revised. If major changes are needed, the Grade 3 Pretest Test Specifications Committee will meet again.
The final test specifications will be available on the web.

Test Specifications
The North Carolina End-of-Grade Test of Mathematics—Grade 3
Aligned to the 2003 Mathematics Curriculum

1. Timeline for Aligning End-of-Grade—Grade 3 Tests to the 2003 Mathematics Curriculum

Year	Activity
2003–2004	Test Specifications, Item Development
2004–2005	Stand-alone Field Test (paper/pencil and online for use in NCCATS)
2005–2006	1st year of Operational Test (with embedded field test items)

2. Purpose and Uses of the Test
The North Carolina End-of-Grade Test of Mathematics—Grade 3 (Mathematics End-of-Grade—Grade 3) is one component of the end-of-grade tests, which include reading comprehension and mathematics tests in grades three through eight. End-of-Grade 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.*"

Student scores will be used in determining student progress and proficiency under state-mandated statewide Student Accountability Standards at grade 3. 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 Mathematics End-of-Grade—Grade 3 test.

Student scores on the Mathematics End-of-Grade—Grade 3 test will also be used in the computation of school growth and performance composites as required by the state mandated ABCs Accountability Program and for determining adequate yearly progress (AYP) under Title I mandates of the No Child Left Behind Act of 2001.

3. Eligible Students
All students in membership in grade 3 shall participate in the North Carolina Statewide Testing Program. Students are expected to participate in the administration of the end-of-grade test in mathematics. The only exceptions are as follows (although an answer sheet must be coded for every student in membership):

- Students with disabilities who participate in the North Carolina Alternate Assessment Portfolio in accordance with state policies; and

- Students who participate in the North Carolina Alternate Assessment Academic Inventory for Mathematics in accordance with state policies (are not administered the end-of-grade mathematics test).

4. Grade 3-Level Competencies in the 2003 Mathematics *Standard Course of Study*

Major Concepts/Skills	**Concepts/Skills to Maintain**
Number sense 0–9,999	Addition and subtraction of multidigit numbers
Multiplication and division	Length and time
Non-negative rational numbers	Symmetry and congruence
Capacity and mass	Line plots, tallies, pictographs
Coordinate grids	Venn diagrams
Circle graphs	
Permutations and combinations	
Growing patterns	
Variables	
Students will solve relevant and authentic problems using appropriate technology and apply these concepts as well as those developed in earlier years	

Strands: Number and Operations, Measurement, Geometry, Data Analysis and Probability, Algebra

COMPETENCY GOAL 1 (Number and Operations): The learner will model, identify, and compute with whole numbers through 9,999.

Objectives

1.01 Develop number sense for whole numbers through 9,999.
- Connect model, number word, and number using a variety of representations.
- Build understanding of place value (ones through thousands).
- Compare and order.

1.02 Develop fluency with multidigit addition and subtraction through 9,999 using:
- Strategies for adding and subtracting numbers.
- Estimation of sums and differences in appropriate situations.
- Relationships between operations.

1.03 Develop fluency with multiplication from 1x1 to 12x12 and division up to two-digit by one-digit numbers using:
- Strategies for multiplying and dividing numbers.
- Estimation of products and quotients in appropriate situations.
- Relationships between operations.

1.04 Use basic properties (identity, commutative, associative, order of operations) for addition, subtraction, multiplication, and division.

1.05 Use area or region models and set models of fractions to explore part-whole relationships.
- Represent fractions concretely and symbolically (halves, fourths, thirds, sixths, eighths).

- Compare and order fractions (halves, fourths, thirds, sixths, eighths) using models and benchmark numbers (zero, one-half, one); describe comparisons.
- Model and describe common equivalents, especially relationships among halves, fourths and eighths, and thirds and sixths.
- Understand that the fractional relationships that occur between zero and one also occur between every two consecutive whole numbers.
- Understand and use mixed numbers and their equivalent fraction forms.

1.06 Develop flexibility in solving problems by selecting strategies and using mental computation, estimation, calculators or computers, and paper and pencil.

COMPETENCY GOAL 2 (Measurement): The learner will recognize and use standard units of metric and customary measurement.

Objectives
2.01 Solve problems using measurement concepts and procedures involving:
- Elapsed time.
- Equivalent measures within the same measurement system.

2.02 Estimate and measure using appropriate units.
- Capacity (cups, pints, quarts, gallons, liters). Length (miles, kilometers).
- Mass (ounces, pounds, grams, kilograms).
- Temperature (Fahrenheit, Celsius).

COMPETENCY GOAL 3 (Geometry): The learner will recognize and use basic geometric properties of two- and three-dimensional figures.

Objectives
3.01 Use appropriate vocabulary to compare, describe, and classify two- and three-dimensional figures.
3.02 Use a rectangular coordinate system to solve problems.
- Graph and identify points with whole number and/or letter coordinates.
- Describe the path between given points on the plane.

COMPETENCY GOAL 4 (Data Analysis and Probability): The learner will understand and use data and simple probability concepts.

Objectives
4.01 Collect, organize, analyze, and display data (including circle graphs and tables) to solve problems.
4.02 Determine the number of permutations and combinations of up to three items.
4.03 Solve probability problems using permutations and combinations.

COMPETENCY GOAL 5 (Algebra): The learner will recognize, determine, and represent patterns and simple mathematical relationships.

Objectives
5.01 Describe and extend numeric and geometric patterns.
5.02 Extend and find missing terms of repeating and growing patterns.

5.03 Use symbols to represent unknown quantities in number sentences.
5.04 Find the value of the unknown in a number sentence.

5. Objectives Measured Indirectly

Some objectives are pedagogical in nature and are not intended to be measured on a multiple-choice test. Items will not be written to specifically measure these objectives. However, they are measured indirectly because they need to be mastered before a student is able to fully master other objectives or grade-level competencies.

Some curricular objectives are written in broad terms and are intended to cover a variety of competencies that can be found in other objectives. An item on the Mathematics End-of-Grade—Grade 3 test must be matched to one and only one objective. Therefore, for item-development purposes, items will not be written to specifically measure objectives that are written in broad terms. However, they are tested indirectly.

Goal and Strand	Objective(s) Measured Indirectly	Remarks
1: Number and Operations	1.06	Includes competencies that appear in other objectives
2: Measurement	none	
3: Geometry	none	
4: Data Analysis & Probability	none	
5: Algebra	5.03	Pedagogical and is needed for 5.04

6. Miscellaneous Remarks about Objectives

The following table contains remarks made about some of the grade 3 objectives. These remarks are listed to remind members of the Mathematics End-of-Grade—Grade 3 Test Specifications Committee of discussions about the objectives and to assist with item development. The table is not intended to direct instruction. For information about the objectives, review the *Grade Three Indicators for the Mathematics* Standard Course of Study. The Grade 3 Indicators are available at www.learnnc.org/dpi/instserv.nsf/category7.

Objective	Remarks
1.02c	Only use addition and subtraction operations
1.03a	Students expected to do this without the calculator
1.03b	Students will be taught a variety of methods for estimating; students are expected to do this without the calculator
1.03c	Students expected to do this without the calculator; only use multiplication and division operations
1.04	Focus is on using the property; students should understand when a property is used properly; no questions on identifying the property; order of operations should be done without a calculator; may have order-of-operation problems involving parentheses
1.05b	Compare and order fractions (can compare mixed numbers such as 4 ½ and 4 ¼) ; does not require common denominators; modeling is essential; within an item always use the same unit or whole; use 0, ½, and 1 as benchmarks (Example: Which is closest to 1?)
3.01	Vocabulary is found in the Grade 3 Indicators (available at www.learnnc.org/dpi/instserv.nsf/category7)
4.01	Try to include data from science or social studies; can use any representations that are covered in K–3
Overall	Students will be expected to work with money. However, they will not be expected to combine dollars and cents using decimal notation (e.g. use 535¢, 535 cents, but not $5.35; use $5 or 5 dollars, but not $5.00)

7. Percent of Items by Goal per Test Form
The test will be designed to meet a percent-by-goal specification rather than a number-per-objective specification.

Goal and Strand	Percent	Priority of Objective Coverage
1: Number & Operations	35–40%	1.03, 1.05, 1.02, 1.01, 1.04
2: Measurement	10–12%	2.02, 2.01
3: Geometry	12–15%	3.01, 3.02
4: Data Analysis & Probability	12–15%	4.01, 4.02, 4.03
5: Algebra	20–25%	equal priority

8. Reporting by Strand
After the operational test is administered and standards have been set, Goal Summary Reports will be available for schools. The Goal Summary Report will provide information by goal/strand. Information about a strand can be reported if a minimum of 6 items for that strand is on the test. If needed, the Measurement and Geometry strands will be combined in the report.

9. Design of Assessments Aligned to the 2003 Mathematics Curriculum
Mathematics End-of-Grade—Grade 3 will be designed to be accessible to as many students as possible to reduce the need for the use of accommodations. Item writers will be trained on the universal design principles to ensure that items are accessible to Exceptional Children and Limited English Proficient students. The NCDPI consultants representing the needs of Exceptional Children and Limited English Proficient students will be expected to work with NCDPI Testing and Curriculum staff in reviewing and editing test items. Special consideration will be given to how items will be read aloud to students requiring a "read-aloud" accommodation and for how items will be signed to hearing-impaired students.

Items must be accessible to students taking the test using regular-print test booklets, large-print test booklets, Braille editions, or a computer screen. **Items that cannot be adapted to fit into any one or more of these modes will not be used on the assessment**.

10. Mode of Presentation
Test items will be in multiple-choice format with only one correct answer choice and three plausible distractors. Items will be designed to fit into the following modes of presentation:

- regular-print test booklets
- large-print test booklets
- Braille editions
- computer screens

11. Some Technical Characteristics of Items

- Distractors should all be plausible (e.g., common errors or misconceptions) so that they cannot easily be ruled out.
- Words should not get in the way of the construct that is being measured.
- Students should not be able to get the correct answer by doing the wrong mathematics operations/processes. For example, What is 2 × 2 ? A student may answer 4 because 2+2 is 4.

12. Levels of Difficulty

Item writers and reviewers will be asked to classify items by level of difficulty. This is done to assist item writers in developing items that measure a broad range of abilities. Prior to field testing, North Carolina educators will use professional judgments to place an item into one of three categories: Easy, Medium, or Hard. These professional judgments will also be used to guide item placement in field tests.

Level of Difficulty	Percent of Items (per form)
Easy	25%
Medium	50%
Hard	25%

The average p-value (the percent of students getting the item correct) of all the items on the test should preferably be in the vicinity of .625. However, as a minimum, the overall p-value will be the average of the operational item pool.

13. Thinking Skills

The thinking skills framework used in developing Mathematics End-of-Grade—Grade 3 is from *Dimensions of Thinking* by Robert J. Marzano and others (1988). Items will be categorized as requiring the use of lower-order or higher-order thinking skills.

Lower-Order Thinking Skills	Higher-Order Thinking Skills
Knowledge	Analyzing
Organizing	Generating
Applying	Integrating
	Evaluating
40% per form (maximum)	60% per form (minimum)

14. Number of Items per Form

Each test form will consist of 82 multiple-choice items. Of these, 50 will be operational items and 32 will be field test items. Operational items will be used to calculate a student's overall test score. Field test items will not be used to calculate a student's test score. Items will be placed in either the Calculator Active or Calculator Inactive part of the test.

	Operational Items	Field test Items
Calculator Active	36	18
Calculator Inactive	14	14
Total	50	32

15. Test Format
The test will consist of two parts as shown in the table below.

Calculator Active Part	Calculator Inactive Part
Calculator allowed	No calculator allowed
Approximately 70% of the test	Approximately 30% of the test
	Must include 1.02a, 1.02b, 1.03a, 1.03b, 1.03c, and order of operations from 1.04
3 sections (18 items each)	2 sections (14 items each)
Graph paper and blank paper will be provided	Graph paper and blank paper will be provided

16. Limited English Proficient Students and Testing Accommodations
On a case-by-case basis where appropriate documentation exists, students identified as limited English proficient may receive testing accommodations. The need for accommodations must be documented. For information regarding the appropriate testing procedures, test administrators who provide accommodations for students identified as limited English proficient must refer to the most recent publication of *Guidelines for Testing Students with Limited English Proficiency* and any published supplements or updates. The procedures for providing accommodations take precedence over other information located in the *Test Administrator's Manual*. Test administrators conducting test administrations with accommodations must be trained in the use of the specified accommodations by the school system test coordinator or designee prior to the test administration.

17. Students with Disabilities and Testing Accommodations
On a case-by-case basis where appropriate documentation exists, students with disabilities, including students **only** receiving services under Section 504, may receive testing accommodations. The need for accommodations for students must be documented in the student's current IEP or Section 504 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* (published 2003) and any published supplements or updates. This publication is available through the local school system or at www.ncpublicschools.org/accountability/testing. The procedures for providing accommodations in these publications take precedence over other information located in the *Test Administrator's Manual*. Test administrators conducting test administrations with accommodations must be trained in the use of the specified accommodations by the school system test coordinator or designee prior to the test administration. In order to provide students with disabilities with modified test formats (i.e., Braille, large print, or One Test Item Per Page editions) as testing accommodations, the school system test coordinator must specifically order such materials at least 60 days prior to the test administration.

18. Accommodations
Below are the typical accommodations available for the End-of-Grade Test of Mathematics—Grade 3 that will result in a valid administration.

Typical Accommodations for End-of-Grade Test of Mathematics—Grade 3*

Accommodation:	**Students with Disabilities/ Section 504**	**Students Identified as Limited English Proficient**
Assistive Technology Devices***	Yes	No
Braille Edition	Yes	No
Braille Writer/Slate and Stylus (Braille Paper)	Yes	No
Cranmer Abacus	Yes	No
Dictation to a Scribe	Yes	No
English/Native Language Dictionary or Electronic Translator	No	Yes
Home/Hospital Testing	Yes	No
Interpreter/Transliterator Signs/Cues Test	Yes	No
Keyboarding Devices	Yes	No
Large Print Edition	Yes	No
Magnification Devices	Yes	No
Multiple Testing Sessions	Yes	Yes
One Test Item Per Page	Yes	No
Scheduled Extended Time	Yes	Yes
Student Marks Answers in Test Book	Yes	Yes
Test Administrator Reads Test Aloud (in English)	Yes	Yes
Testing in a Separate Room	Yes	Yes

*School system personnel must participate in training sessions prior to the test administration to be aware of any restrictions for appropriate use of these accommodations.
NOTE: The *Testing Students with Disabilities* document (published February 2003) provides additional information on accommodations and guidelines for testing students with disabilities.

19. Testing Window
School systems shall direct schools to administer end-of-grade tests on consecutive school days during the last three weeks of the school year. For school systems that were required to adjust their school schedules due to adverse weather conditions and other emergencies, the testing schedule is to be adjusted to fall within the final three weeks of the adjusted school year.

20. Test Schedule
The tests must be administered as early in the school day as the school schedule permits. Afternoon administrations of the end-of-grade tests are prohibited. All students at the same grade level within a school must be administered the appropriate end-of-grade test at the same time on the same day.

21. Estimated Administration Time
Students are expected to complete state tests in the allotted test-administration time (unless a student with a disability or limited English proficiency has the appropriate documentation in accordance with state policies to receive time-modified accommodations).

The times noted below are estimations:

DAY 1

End-of-Grade Test of Mathematics—Grade 3	Items	Minutes
Calculator Active Part		
General Instructions		12
Section 1	18	40*
Break		3
Section 2	18	40*
Break		3
Section 3	18	40*
Estimated Total for Calculator Active Part Only	**54**	**138**

DAY 2

Calculator Inactive Part	Items	Minutes
General Instructions		10
Section 1	14	30*
Break		3
Section 2	14	30*
Estimated Total for Calculator Inactive Part Only	**28**	**73**

*Each section is timed. After the time allotted for the section has expired, a student will not be allowed to return to the section.

The mathematics test must be administered on two separate consecutive school days. The calculator active part of the mathematics test must be administered on Day 1. The calculator inactive part must be administered on Day 2.

22. Additional Assessment Materials
Each student will be given graph paper and blank paper at the beginning of each test administration. A calculator is allowed on the Calculator Active part of the test only. No formula sheets will be required. No manipulatives (e.g. rulers, protractors, pattern blocks) will be required.

23. Calculator Use
All students must have access to calculators during the administration of the calculator active part of the test. However, students may NOT use calculators during the administration of the calculator inactive part of the test. Student use of calculators during the calculator inactive sections of the test results in a misadministration.

The minimum calculator requirements listed below denote what every calculator must have.

Minimum ("at least") Calculator Requirement
Calculator Inactive: Calculator use is not allowed.
Calculator Active: Four-function calculator with memory key

Additional features that are not restricted (see below) are **allowed but are not required**.

*****Restrictions*****
Students are not allowed to use calculators with a typewriter-style (QWERTY) keyboard or calculators that include a computer algebraic system (CAS) and are capable of doing symbolic algebra (i.e., factoring, expanding, or simplifying given variable output), *or symbolic calculus. Cell phones, handheld microcomputers, pen-input devices* (such as personal digital assistants or tablets), *or laptop/notebook computers are not allowed during any of the test administrations.* As curricula and technology change, the policy concerning technology use on North Carolina tests will be reevaluated.

Students are not allowed to share calculators during the test administration.

Students, who regularly use more than one calculator during classroom instructional activities, may be permitted to use more than one calculator during the test administration. Students may use calculators with fraction keys; however, the use of fraction keys without prior training may confuse students and may adversely affect their performance during the test administration.

24. Scoring and Reporting
The school system test coordinator establishes the schedule for scanning and scoring the end-of-grade answer sheets. Scanning, scoring, and initial district- and school-level reporting timelines are established locally. The school system test coordinator will provide the results (reports) from the test administration soon after scanning/scoring is completed. The NCDPI will provide descriptive information about the average scores obtained by all students and specific groups of students by school, the school system, and the state at the time the school accountability results are presented to the State Board of Education.

25. Special Requests
The following special requests made by the Curriculum, EC, or LEP staff during various meetings will be honored:

- "Not" and "except" questions will be avoided whenever possible.
- On the sample page, there will be written instructions to "Choose the ***best*** answer."

- On the sample page, students will have a written reminder that “Not all diagrams are drawn to scale.”
- Starting at grade 4, the sample page will contain a statement informing students that some fractions are simplified.
- There will be a different sample question for grades 3–5 and grades 6–8.
- “About” and “Estimate” can be used at all grades. “Approximately” will not be used in grades 3–5.

Guidelines for Item Development and a Style Manual are available for guiding item development. In addition, the NCDPI consultants representing the needs of Exceptional Children and Limited English Proficient students have written documents outlining “Things to Consider” when writing test items. At the start of each testing year, these documents are updated.

26. Meetings and Attendance
Preliminary meetings were held to discuss how to align tests to the 2003 curriculum. Decisions made at these meetings helped to guide discussions at the Mathematics End-of-Grade—Grade 3 Test Specifications meeting held on July 23 and 24, 2003.

Preliminary Meetings to Discuss Aligning Tests to the 2003 Math Curriculum
April 11, 2003, room 694, Education Building, 9 a.m.–11 a.m.
Attendance
Sarah McManus, Lead Consultant for Mathematics Assessments, NCDPI Testing
Laura Kramer, Senior Psychometrician, NCDPI Testing
Bill Tucci, Section Chief for Mathematics and Science, NCDPI Instructional Services
Toni Meyer, Mathematics Consultant, K–5, NCDPI Instructional Services
Bill Scott, Mathematics Consultant, 9–12, NCDPI Instructional Services
Linda Patch, Mathematics Consultant, 6–8, NCDPI Instructional Services

April 30, 2003, 5th floor library, Education Building, 11 a.m.–12:30 p.m.
Attendance
Sarah McManus, Lead Consultant for Mathematics Assessments, NCDPI Testing
Kevin Murphy, Operations Manager, NCDPI Testing
Toni Meyer, Mathematics Consultant, K–5, NCDPI Instructional Services
Bill Scott, Mathematics Consultant, 9–12, NCDPI Instructional Services
Linda Patch, Mathematics Consultant, 6–8, NCDPI Instructional Services
Diann Irwin, Section Chief for Behavior Support Services, NCDPI Exceptional Children
Martha Downing, Consultant for the Hearing Impaired, NCDPI, Exceptional Children
Tom Winton, Consultant for the Visual Impaired, NCDPI, Exceptional Children

May 15, 2003, Room 150, Education Building, 2 p.m.–4 p.m.
Attendance
Lou Fabrizio, Director, Division of Accountability Services, NCDPI
Mildred Bazemore, Section Chief for Test Development, NCDPI
Sarah McManus, Lead Consultant for Mathematics Assessments, NCDPI Testing
Laura Kramer, Senior Psychometrician, NCDPI Testing
June Atkinson, Director, Division of Instructional Services, NCDPI
Frances Hoch, Section Chief for Second Languages, ESL, Information and Computer Skills,

NCDPI
Bill Tucci, Section Chief for Mathematics and Science, NCDPI
Toni Meyer, Mathematics Consultant, K–5, NCDPI Instructional Services
Bill Scott, Mathematics Consultant, 9–12, NCDPI Instructional Services
Linda Patch, Mathematics Consultant, 6–8, NCDPI Instructional Services

Test Coordinator's Advisory Committee
May 1, 2003, Room 224, Education Building, 10:45 a.m.–11:15 a.m.
Sarah McManus presented information to the committee about proposed new item formats. Feedback was gathered.

Technical Advisory Committee Meeting
May 8, 2003, SBE lounge, 7th floor, Education Building, 1 p.m.–2 p.m.
Sarah McManus presented information to the committee about proposed new item formats. Feedback was gathered.

Mathematics End-of-Grade—Grade 3 Test Specifications Committee Meeting
July 23, 2003, Room 694, Education Building 9:00 a.m.–4 p.m.
July 24, 2003, Room 694, Education Building 9:00 a.m.–11 a.m.
Attendance
Sarah McManus, Lead Consultant for Mathematics Assessments, NCDPI Testing
Toni Meyer, Mathematics Consultant, K–5, NCDPI Instructional Services
Linda Patch, Mathematics Consultant, 6–8, NCDPI Instructional Services
Donna Taylor, Mathematics Consultant, 6–8, NCDPI Instructional Services
Sallie Abbas, Mathematics Consultant, Technical Outreach for Public Schools
Thomas Englehart, Mathematics Consultant, Technical Outreach for Public Schools
Beth Spivey, Coordinating Teacher, Wake County Public Schools
Johnny Warrick, Mathematics Specialist, K–5, Gaston County Public Schools
Russell Hinson, Lead Teacher, Nathaniel Alexander Elementary, Charlotte
Joyce Hodges, Mathematics Curriculum Specialist, Cumberland County Public Schools

NCDPI EOG Mathematics Test Specifications Meeting
September 4, 2003, Room 224, Education Building, 12:30–2:30 p.m.
Attendance (all are members of the NCDPI staff)
Zoe Locklear, Associate Superintendent, Leadership Development and Social Services
Lou Fabrizio, Director, Division of Accountability Services
June Atkinson, Director, Division of Instructional Services
Mary Watson, Director, Division of Accountability Services
Mildred Bazemore, Section Chief for Test Development
Tammy Howard, Section Chief for Testing Policies and Operations
Bill Tucci, Section Chief for Mathematics and Science
Frances Hoch, Section Chief for Second Languages, ESL, Information and Computer Skills, Instructional Services
David Mills, Section Chief for Areas of Exceptionality, Exceptional Children
Sarah McManus, Lead Consultant for Mathematics Assessments, Testing
Laura Kramer, Senior Psychometrician, Testing
Kelly Burling, Psychometrician, Testing
Pam VanDyk, Technical Writer and Lead Consultant for NCCATS Project, Testing

Toni Meyer, Mathematics Consultant, K–5, Instructional Services
Linda Patch, Mathematics Consultant, 6–8, Instructional Services
Donna Taylor, Mathematics Consultant, 6–8, Instructional Services
Tom Winton, Consultant for the Visual Impaired, Exceptional Children

As needed, this document will be revised. If major changes are needed, the Mathematics End-of-Grade—Grade 3 Test Specifications Committee will meet again.

Test Specifications
The North Carolina End-of-Grade Test of Mathematics—Grade 4
Aligned to the 2003 Mathematics Curriculum

1. Timeline for Aligning End-of-Grade—Grade 4 Tests to the 2003 Mathematics Curriculum [Omitted, as in Grade 3]
2. Purpose and Uses of the Test [Omitted, as in Grade 3]
3. Eligible Students [Omitted, as in Grade 3]

4. Grade 4-Level Competencies in the 2003 Mathematics *Standard Course of Study*

Major Concepts/Skills	Concepts/Skills to Maintain
Number sense 0.01–99,999	Whole number computation
Multiplication and division of multidigit numbers	Non-negative rational numbers
Perimeter and area	Length, time, capacity, and mass
Transformations	Symmetry and congruence
Line graphs	Coordinate grids
Median, mode, and range	Circle graphs
Variables in number sentences	Permutations and combinations
Proportional reasoning	
Students will solve relevant and authentic problems using appropriate technology and apply these concepts as well as those developed in earlier years.	

Strands: Number and Operations, Measurement, Geometry, Data Analysis and Probability, Algebra

COMPETENCY GOAL 1 (Number and Operations): The learner will read, write, model, and compute with non-negative rational numbers.

Objectives
1.01 Develop number sense for rational numbers 0.01 through 99,999.
- Connect model, number word, and number using a variety of representations.
- Build understanding of place value (hundredths through ten thousands).
- Compare and order rational numbers.
- Make estimates of rational numbers in appropriate situations.

1.02 Develop fluency with multiplication and division:
- Two-digit by two-digit multiplication (larger numbers with calculator).
- Up to three-digit by two-digit division (larger numbers with calculator).
- Strategies for multiplying and dividing numbers.
- Estimation of products and quotients in appropriate situations.
- Relationships between operations.

1.03 Solve problems using models, diagrams, and reasoning about fractions and relationships among fractions involving halves, fourths, eighths, thirds, sixths, twelfths, fifths, tenths, hundredths, and mixed numbers.

1.04 Develop fluency with addition and subtraction of non-negative rational numbers with like denominators, including decimal fractions through hundredths.

- Develop and analyze strategies for adding and subtracting numbers.
- Estimate sums and differences.
- Judge the reasonableness of solutions.

1.05 Develop flexibility in solving problems by selecting strategies and using mental computation, estimation, calculators or computers, and paper and pencil.

COMPETENCY GOAL 2 (Measurement): The learner will understand and use perimeter and area.

Objectives

2.01 Develop strategies to determine the area of rectangles and the perimeter of plane figures.
2.02 Solve problems involving perimeter of plane figures and areas of rectangles.

COMPETENCY GOAL 3 (Geometry): The learner will recognize and use geometric properties and relationships.

Objectives

3.01 Use the coordinate system to describe the location and relative position of points and draw figures in the first quadrant.
3.02 Describe the relative position of lines using concepts of parallelism and perpendicularity.
3.03 Identify, predict, and describe the results of transformations of plane figures.

- Reflections.
- Translations.
- Rotations.

COMPETENCY GOAL 4 (Data Analysis and Probability): The learner will understand and use graphs, probability, and data analysis.

Objectives

4.01 Collect, organize, analyze, and display data (including line graphs and bar graphs to solve problems.
4.02 Describe the distribution of data using median, range and mode.
4.03 Solve problems by comparing two sets of related data.
4.04 Design experiments and list all possible outcomes and probabilities for an event.

COMPETENCY GOAL 5 (Algebra): The learner will demonstrate an understanding of mathematical relationships.

Objectives

5.01 Identify, describe, and generalize relationships in which:

- Quantities change proportionally.
- Change in one quantity relates to change in a second quantity.

5.02 Translate among symbolic, numeric, verbal, and pictorial representations of number relationships.
5.03 Verify mathematical relationships using:

- Models, words, and numbers.

- Order of operations and the identity, commutative, associative, and distributive properties.

5. Objectives Measured Indirectly

Some objectives are pedagogical in nature and are not intended to be measured on a multiple-choice test. Items will not be written to specifically measure these objectives. However, they are measured indirectly because they need to be mastered before a student is able to fully master other objectives or grade-level competencies.

Some curricular objectives are written in broad terms and are intended to cover a variety of competencies that can be found in other objectives. An item on the Mathematics End-of-Grade—Grade 4 test must be matched to one and only one objective. Therefore, for item-development purposes, items will not be written to specifically measure objectives written in broad terms. However, they are tested indirectly.

Goal and Strand	Objective(s) Measured Indirectly	Reason(s)
1: Number & Operations	1.02c, 1.04c, 1.05	Pedagogical and are needed for mastering the competencies found in other objectives
2: Measurement	2.01	Pedagogical and is needed for mastering competencies found in other objectives (e.g. 2.02)
3: Geometry	none	
4: Data Analysis & Probability	none	
5: Algebra	none	

6. Miscellaneous Remarks about Objectives

The following table contains remarks made about some of the grade 4 objectives. These remarks are listed to remind members of the Mathematics End-of-Grade—Grade 4 Test Specifications Committee of discussions about the objectives and to assist with item development. The table is not intended to direct instruction. For more information about the objectives, review the *Grade Four Indicators for the Mathematics* Standard Course of Study. The Grade 4 Indicators are available at www.learnnc.org/dpi/instserv.nsf/category7.

Objective	**Remarks**
1.02a	Students can use any strategy; division can be with or without remainders; make sure there are a variety of answers to questions (include quotients, remainders, and dividends)
1.03	Must have diagrams; rectangular models are easier to see than circular ones
2.02	Not just formulas; use grids sparingly; can use figures that can be split into rectangles
3.01	Involves using ordered pairs
3.02	Only use lines in a plane; no skew lines
3.03	Emphasize vocabulary by using “reflection,” “translation,” and “rotation”; coordinate plane should not be used for transformations
4.01	This does not include multiple representations of the same data; can use any method of representation that is covered in K–4
4.03	Be sure to use “related” data not the “same data”; can include a double bar graph
5.01	Not “a is to b as d is to ___;” simple rate and time problems; function machine
5.02	Use balance scales as well as other representations
5.03a	Students should be able to justify steps in a problem
5.03b	Make sure students using calculators that have algebraic logic do not have an advantage; use symbols

7. Percent of Items by Goal per Test Form
The test will be designed to meet a percent-by-strand specification rather than a number-per-objective specification.

Goal and Strand	Percent	Priority of Objective Coverage
1: Number & Operations	35–40%	1.03, 1.04, 1.01 (emphasis is on decimals), 1.02
2: Measurement	10–12%	Not applicable
3: Geometry	10–12%	3.02, 3.03, 3.01
4: Data Analysis & Probability	15–18%	4.02, 4.01, 4.04, 4.03
5: Algebra	20–25%	5.01, 5.02, 5.03

8. Reporting by Strand [Omitted, as in Grade 3]
9. Design of Assessments Aligned to the 2003 Mathematics Curriculum [Omitted, as in Grade 3]
10. Mode of Presentation [Omitted, as in Grade 3]
11. Some Technical Characteristics of Items [Omitted, as in Grade 3]
12. Levels of Difficulty [Omitted, as in Grade 3]
13. Thinking Skills [Omitted, as in Grade 3]
14. Number of Items per Form [Omitted, as in Grade 3]

15. Test Format
The test will consist of two parts as shown in the table below.

Calculator Active Part	Calculator Inactive Part
Calculator allowed	No calculator allowed
Approximately 70% of the test	Approximately 30% of the test
Must include multiplication of numbers larger than 2-digit by 2-digit Must include division involving larger than 3-digit numbers divided by 2-digit numbers	Must include 1.02a, 1.02b, 1.04a, 1.04b
3 sections (18 items each)	2 sections (14 items each)
Graph paper and blank paper will be provided	Graph paper and blank paper will be provided

16. Limited English Proficient Students and Testing Accommodations [Omitted, as in Grade 3]
17. Students with Disabilities and Testing Accommodations [Omitted, as in Grade 3]
18. Accommodations [Omitted, as in Grade 3]
19. Testing Window [Omitted, as in Grade 3]
20. Test Schedule [Omitted, as in Grade 3]

21. Estimated Administration Time [Omitted, as in Grade 3]
22. Additional Assessment Materials [Omitted, as in Grade 3]
23. Calculator Use [Omitted, as in Grade 3]
24. Scoring and Reporting [Omitted, as in Grade 3]
25. Special Requests [Omitted, as in Grade 3]
26. Meetings and Attendance [Omitted, as in Grade 3]

Test Specifications
The North Carolina End-of-Grade Test of Mathematics—Grade 5
Aligned to the 2003 Mathematics Curriculum

1. Timeline for Aligning End-of-Grade—Grade 5 Tests to the 2003 Mathematics Curriculum [Omitted, as in Grade 3]
2. Purpose and Uses of the Test [Omitted, as in Grade 3]
3. Eligible Students [Omitted, as in Grade 3]

4. Grade 5-Level Competencies in the 2003 Mathematics *Standard Course of Study*

Major Concepts/Skills	Concepts/Skills to Maintain
Number sense 0.001–999,999	Whole number computation
Addition and subtraction of non-negative rational numbers	Transformations
Properties of plane figures	Perimeter and area
Bar graphs and stem-and-leaf plots	Coordinate grids
Rates of change	Line graphs
Simple equations and inequalities	Median, mode, and range
Students will solve relevant and authentic problems using appropriate technology and apply these concepts as well as those developed in earlier years	

Strands: Number and Operations, Measurement, Geometry, Data Analysis and Probability, Algebra

COMPETENCY GOAL 1 (Number and Operations): The learner will understand and compute with non-negative rational numbers.

Objectives
1.01 Develop number sense for rational numbers 0.001 through 999,999.
- Connect model, number word, and number using a variety of representations.
- Build understanding of place value (thousandths through hundred thousands).
- Compare and order rational numbers.
- Make estimates of rational numbers in appropriate situations.

1.02 Develop fluency in adding and subtracting non-negative rational numbers (halves, fourths, eighths; thirds, sixths, twelfths; fifths, tenths, hundredths, thousandths; mixed numbers).
- Develop and analyze strategies for adding and subtracting numbers.
- Estimate sums and differences.
- Judge the reasonableness of solutions.

1.03 Develop flexibility in solving problems by selecting strategies and using mental computation, estimation, calculators or computers, and paper and pencil.

COMPETENCY GOAL 2 (Measurement): The learner will recognize and use standard units of metric and customary measurement.

Objectives
2.01 Estimate the measure of an object in one system given the measure of that object in another system.
2.02 Identify, estimate, and measure the angles of plane figures using appropriate tools.

COMPETENCY GOAL 3 (Geometry): The learner will understand and use properties and relationships of plane figures.
Objectives
3.01 Identify, define, describe, and accurately represent triangles, quadrilaterals, and other polygons.
3.02 Make and test conjectures about polygons involving:

- Sum of the measures of interior angles.
- Lengths of sides and diagonals.
- Parallelism and perpendicularity of sides and diagonals.

3.03 Classify plane figures according to types of symmetry (line, rotational).
3.04 Solve problems involving the properties of triangles, quadrilaterals, and other polygons.

- Sum of the measures of interior angles.
- Lengths of sides and diagonals.
- Parallelism and perpendicularity of sides and diagonals.

COMPETENCY GOAL 4 (Data Analysis and Probability): The learner will understand and use graphs and data analysis.

Objectives
4.01 Collect, organize, analyze, and display data (including stem-and-leaf plots) to solve problems.
4.02 Compare and contrast different representations of the same data; discuss the effectiveness of each representation.
4.03 Solve problems with data from a single set or multiple sets of data using median, range, and mode.

COMPETENCY GOAL 5 (Algebra): The learner will demonstrate an understanding of patterns, relationships, and elementary algebraic representation.

Objectives
5.01 Describe, extend, and generalize numeric and geometric patterns using tables, graphs, words, and symbols.
5.02 Use algebraic expressions, patterns, and one-step equations and inequalities to solve problems.
5.03 Identify, describe, and analyze situations with constant or varying rates of change.

5. Objectives Measured Indirectly
Some objectives are pedagogical in nature and are not intended to be measured on a multiple-choice test. Items will not be written to specifically measure these objectives. However, they are measured indirectly because they need to be mastered before a student is able to fully master other objectives or grade-level competencies.

Some curricular objectives are written in broad terms and are intended to cover a variety of competencies that can be found in other objectives. An item on the Mathematics End-of-Grade—Grade 5 test must be matched to one and only one objective. Therefore, for item-development purposes, items will not be written to specifically measure objectives written in broad terms. However, they are tested indirectly.

Goal and Strand	**Objective(s) Measured Indirectly**	**Reason(s)**
1: Number and Operations	1.02c	Pedagogical and is needed for mastering other objectives
2: Measurement	none	
3: Geometry	3.02	Pedagogical and is needed for mastering other objectives
4: Data Analysis & Probability	none	
5: Algebra	none	

6. Miscellaneous Remarks about Objectives
The following table contains remarks made about some of the grade 5 objectives. These remarks are listed to remind members of the Mathematics End-of-Grade—Grade 5 Test Specifications Committee of discussions about the objectives and to assist with item development. The table is not intended to direct instruction. For more information about the objectives review the *Grade Five Indicators for the Mathematics* Standard Course of Study available at www.learnnc.org/dpi/instserv.nsf/category7.

Objective	**Remarks**
1.01d	Can change from one representation to another (e.g. fraction to decimal)
1.02	Within families; families separated by semicolon in the wording of the objective
2.01	Be able to understand two systems but not convert from one system to the other
2.02	Acute, right, obtuse and straight angles
3.01	Can use figures that are regular, irregular, convex or concave; n-sided polygons for n <=10 can be used; students are not responsible for understanding the terms “concave” and “convex”
5.02	Can involve translating phrases into symbolic language

7. Percent of Items by Goal per Test Form
The test will be designed to meet a percent-by-strand specification rather than a number-per-objective specification.

Goal and Strand	Percent	Priority of Objective Coverage
1: Number & Operations	20–25%	1.02, 1.01, 1.03
2: Measurement	10–15%	Equal priority
3: Geometry	25–30%	3.04, 3.01, 3.03
4: Data Analysis & Probability	10–15%	4.01, 4.03, 4.02
5: Algebra	20–25%	5.02, 5.03, 5.01

8. Reporting by Strand [Omitted, as in Grade 3]
9. Design of Assessments Aligned to the 2003 Mathematics Curriculum [Omitted, as in Grade 3]
10. Mode of Presentation [Omitted, as in Grade 3]
11. Some Technical Characteristics of Items [Omitted, as in Grade 3]
12. Levels of Difficulty [Omitted, as in Grade 3]
13. Thinking Skills [Omitted, as in Grade 3]
14. Number of Items per Form [Omitted, as in Grade 3]

15. Test Format
The test will consist of two parts as shown in the table below.

Calculator Active Part	Calculator Inactive Part
Calculator allowed	No calculator allowed
Approximately 70% of the test	Approximately 30% of the test
Must include multiplication of numbers larger than 2-digit by 2-digit Must include division involving larger than 3-digit numbers divided by 2-digit numbers	Must include 1.02a, 1.02b, 1.01d
3 sections (18 items each)	2 sections (14 items each)
Graph paper and blank paper will be provided	Graph paper and blank paper will be provided

16. Limited English Proficient Students and Testing Accommodations [Omitted, as in Grade 3]
17. Students with Disabilities and Testing Accommodations [Omitted, as in Grade 3]
18. Accommodations [Omitted, as in Grade 3]
19. Testing Window [Omitted, as in Grade 3]

20. Test Schedule [Omitted, as in Grade 3]
21. Estimated Administration Time [Omitted, as in Grade 3]
22. Additional Assessment Materials [Omitted, as in Grade 3]
23. Calculator Use [Omitted, as in Grade 3]
24. Scoring and Reporting [Omitted, as in Grade 3]
25. Special Requests [Omitted, as in Grade 3]
26. Meetings and Attendance [Omitted, as in Grade 3]

Draft Test Specifications
The North Carolina End-of-Grade Test of Mathematics—Grade 6
Aligned to the 2003 Mathematics Curriculum

1. Timeline for Aligning End-of-Grade—Grade 6 Tests to the 2003 Mathematics Curriculum [Omitted, as in Grade 3]
2. Purpose and Uses of the Test [Omitted, as in Grade 3]
3. Eligible Students [Omitted, as in Grade 3]

4. Grade 6-Level Competencies in the 2003 Mathematics *Standard Course of Study*

Major Concepts/Skills	**Concepts/Skills to Maintain**
Negative rational numbers	Addition and subtraction of non-negative rational numbers
Percent	Number properties
Transformations in the coordinate plane	Perimeter and area
Probability	Median, mode, and range
Equations and inequalities	Bar graphs and leaf plots
Multiplication and division of non-negative rational numbers	
Students will solve relevant and authentic problems using appropriate technology and apply these concepts as well as those developed in earlier years	

Strands: Number and Operations, Measurement, Geometry, Data Analysis and Probability, Algebra

COMPETENCY GOAL 1 (Number and Operations): The learner will understand and compute with rational numbers.

Objectives
1.01 Develop number sense for negative rational numbers.
- Connect the model, number word, and number using a variety of representations, including the number line.
- Compare and order.
- Make estimates in appropriate situations.

1.02 Develop meaning for percents.
- Connect the model, number word, and number using a variety of representations.
- Make estimates in appropriate situations.

1.03 Compare and order rational numbers.
1.04 Develop fluency in addition, subtraction, multiplication, and division of non-negative rational numbers.
- Analyze computational strategies.
- Describe the effect of operations on size.
- Estimate the results of computations.
- Judge the reasonableness of solutions.

1.05 Develop fluency in the use of factors, multiples, exponential notation, and prime factorization.
1.06 Use exponential, scientific, and calculator notation to write very large and very small numbers.
1.07 Develop flexibility in solving problems by selecting strategies and using mental computation, estimation, calculators or computers, and paper and pencil.

COMPETENCY GOAL 2 (Measurement): The learner will select and use appropriate tools to measure two- and three-dimensional figures.

Objectives
2.01 Estimate and measure length, perimeter, area, angles, weight, and mass of two- and three-dimensional figures, using appropriate tools.
2.02 Solve problems involving perimeter/circumference and area of plane figures.

COMPETENCY GOAL 3 (Geometry): The learner will understand and use properties and relationships of geometric figures in the coordinate plane.

Objectives
3.01 Identify and describe the intersection of figures in a plane.
3.02 Identify the radius, diameter, chord, center, and circumference of a circle; determine the relationships among them.
3.03 Transform figures in the coordinate plane and describe the transformation.
3.04 Solve problems involving geometric figures in the coordinate plane.

COMPETENCY GOAL 4 (Data Analysis and Probability): The learner will understand and determine probabilities.

Objectives
4.01 Develop fluency with counting strategies to determine the sample space for an event. Include lists, tree diagrams, frequency distribution tables, permutations, combinations, and the Fundamental Counting Principle.
4.02 Use a sample space to determine the probability of an event.
4.03 Conduct experiments involving simple and compound events.
4.04 Determine and compare experimental and theoretical probabilities for simple and compound events.
4.05 Determine and compare experimental and theoretical probabilities for independent and dependent events.
4.06 Design and conduct experiments or surveys to solve problems; report and analyze results.

COMPETENCY GOAL 5 (Geometry): The learner will demonstrate an understanding of simple algebraic expressions.

Objectives
5.01 Simplify algebraic expressions and verify the results using the basic properties of rational numbers.

- Identity.
- Commutative.

- Associative.
- Distributive.
- Order of operations.

5.02 Use and evaluate algebraic expressions.
5.03 Solve simple (one- and two-step) equations or inequalities.
5.04 Use graphs, tables, and symbols to model and solve problems involving rates of change and ratios.

5. Objectives Measured Indirectly
Some objectives are pedagogical in nature and are not intended to be measured on a multiple-choice test. Items will not be written to specifically measure these objectives. However, they are measured indirectly because they need to be mastered before a student is able to fully master other objectives or grade-level competencies.

Some curricular objectives are written in broad terms and are intended to cover a variety of competencies that can be found in other objectives. An item on the Mathematics End-of-Grade—Grade 6 test must be matched to one and only one objective. Therefore, for item-development purposes, items will not be written to specifically measure objectives written in broad terms. However, they are tested indirectly.

Goal and Strand	Objective(s) Measured Indirectly	Reason(s)
1: Number and Operations	1.01b, 1.01c, 1.07	Competencies are included in other objectives
2: Measurement	none	
3: Geometry	none	
4: Data Analysis & Probability	4.02 4.03	Competencies are included in other objectives (4.05, 4.04) Pedagogical
5: Algebra	none	

6. Miscellaneous Remarks about Objectives
The following table contains remarks made about some of the grade 6 objectives. These remarks are listed to remind members of the Mathematics End-of-Grade—Grade 6 Test Specifications Committee of discussions about the objectives and to assist with item development. The table is not intended to direct instruction. For instructional strategies and more information about the objectives, one should review the *Grade Six Indicators for the Mathematics* Standard Course of Study available at www.learnnc.org/dpi/instserv.nsf/Category7.

Objective	Remarks
1.04	Match items to the whole objective
1.06	Does not include computing
3.01	Students are not expected to know the special names of angles formed by lines cut by a transversal (e.g. alternate interior angles)
3.03	All quadrants can be used; does not include dilations
4.06	Emphasize “report and analyze” experiments or surveys
5.01	Match items to the whole objective

7. Percent of Items by Goal per Test Form
The test will be designed to meet a percent-by-strand specification rather than a number-per-objective specification.

Goal and Strand	Percent	Priority of Objective Coverage
1: Number & Operation	20–25%	1.04, 1.05, 1.01, 1.02, 1.03, 1.06
2: Measurement	10–15%	2.02, 2.01
3: Geometry	15–20%	3.03, 3.02, 3.01, 3.04
4: Data Analysis & Probability	20–25%	4.04, 4.05, 4.01, 4.06
5: Algebra	20–25%	5.02, 5.03, 5.04, 5.01

8. **Reporting by Strand** [Omitted, as in Grade 3]
9. **Design of Assessments Aligned to the 2003 Mathematics Curriculum** [Omitted, as in Grade 3]
10. **Mode of Presentation** [Omitted, as in Grade 3]
11. **Some Technical Characteristics of Items** [Omitted, as in Grade 3]
12. **Levels of Difficulty** [Omitted, as in Grade 3]
13. **Thinking Skills** [Omitted, as in Grade 3]

14. Number of Items per Form
Each test form will consist of 80 multiple-choice items. Of these, 49 will be operational items and 31 will be field test items. Operational items will be used to calculate a student's overall test score. Field test items will not be used to calculate a student's test score. Items will be placed in either the Calculator Active or Calculator Inactive part of the test.

	Operational Items	**Field Test Items**
Calculator Active	36	18
Calculator Inactive	13	13
Total	**49**	**31**

15. Test Format
The test will consist of two parts as shown in the table below.

Calculator Active Part	**Calculator Inactive Part**
Calculator allowed	No calculator allowed
Approximately 70% of the test	Approximately 30% of the test
Must include multiplication of numbers larger than 2-digit by 2-digit Must include division involving larger than 3-digit numbers divided by 2-digit numbers	Must include 1.03, 1.04c
3 sections (18 items each)	2 sections (13 items each)
Graph paper and blank paper will be provided	Graph paper and blank paper will be provided

16. Limited English Proficient Students and Testing Accommodations [Omitted, as in Grade 3]
17. Students with Disabilities and Testing Accommodations [Omitted, as in Grade 3]
18. Accommodations [Omitted, as in Grade 3]
19. Testing Window [Omitted, as in Grade 3]
20. Test Schedule [Omitted, as in Grade 3]

21. Estimated Administration Time
Students are expected to complete state tests in the allotted test-administration time (unless a student with a disability or limited English proficiency has the appropriate documentation in accordance with state policies to receive time-modified accommodations).

The times noted below are estimations:

DAY 1

End-of-Grade Test of Mathematics—Grade 3	Items	Minutes
Calculator Active Part		
General Instructions		12
Section 1	18	40*
Break		3
Section 2	18	40*
Break		3
Section 3	18	40*
Estimated Total for Calculator Active Part Only	**56**	**138**

DAY 2

Calculator Inactive Part	Items	Minutes
General Instructions		10
Section 1	13	30*
Break		3
Section 2	13	30*
Estimated Total for Calculator Inactive Part Only	**26**	**73**

*Each section is timed. After the time allotted for the section has expired, a student will not be allowed to return to the section.

The mathematics test must be administered on two separate consecutive school days. The calculator active part of the mathematics test must be administered on Day 1. The calculator inactive part must be administered on Day 2.

22. Additional Assessment Materials [Omitted, as in Grade 3]

23. Calculator Use
All students must have access to calculators during the administration of the calculator active part of the test. However, students may NOT use calculators during the administration of the calculator inactive part of the test. Student use of calculators during the calculator inactive sections of the test results in a misadministration.

The minimum calculator requirements listed below denote what every calculator must have.

Minimum ("at least") Calculator Requirement
Calculator Inactive: Calculator use is not allowed.
Calculator Active: Any four-function calculator with square root function, y^x, π (*pi*), and algebraic logic

Additional features that are not restricted (see below) are **allowed but are not required**.

*******Restrictions*******
Students are <u>*not*</u> *allowed to use calculators with a typewriter-style (QWERTY) keyboard or calculators that include a computer algebraic system (CAS) and are capable of doing symbolic algebra* (i.e., factoring, expanding, or simplifying given variable output), *or symbolic calculus. Cell phones, handheld microcomputers, pen-input devices (such as personal digital assistants or tablets), or laptop/notebook computers are* <u>*not*</u> *allowed during any of the test administrations*. As curricula and technology change, the policy concerning technology use on North Carolina tests will be reevaluated.

Students are <u>no</u>t allowed to share calculators during the test administration.

Students, who regularly use more than one calculator during classroom instructional activities, <u>may</u> be permitted to use more than one calculator during the test administration. Students may use calculators with fraction keys; however, the use of fraction keys without prior training may confuse students and may adversely affect their performance during the test administration.

24. Scoring and Reporting [Omitted, as in Grade 3]
25. Special Requests [Omitted, as in Grade 3]

26. Meetings and Attendance
Preliminary meetings were held to discuss how to align tests to the 2003 curriculum. Decisions made at these meetings helped to guide discussions at the Mathematics End-of-Grade—Grade 6 Test Specifications meeting held on August 20 and 21, 2003.

Preliminary Meetings to Discuss Aligning Tests to the 2003 Math Curriculum
April 11, 2003, room 694, Education Building, 9 a.m.–11 a.m.
Attendance
Sarah McManus, Lead Consultant for Mathematics Assessments, NCDPI Testing,
Laura Kramer, Senior Psychometrician, NCDPI Testing
Bill Tucci, Section Chief for Mathematics and Science, NCDPI Instructional Services
Toni Meyer, Mathematics Consultant, K–5, NCDPI Instructional Services
Bill Scott, Mathematics Consultant, 9–12, NCDPI Instructional Services
Linda Patch, Mathematics Consultant, 6–8, NCDPI Instructional Services

April 30, 2003, 5th floor library, Education Building, 11 a.m.–12:30 p.m.
Attendance
Sarah McManus, Lead Consultant for Mathematics Assessments, NCDPI Testing,
Kevin Murphy, Operations Manager, NCDPI Testing
Toni Meyer, Mathematics Consultant, K–5, NCDPI Instructional Services
Bill Scott, Mathematics Consultant, 9–12, NCDPI Instructional Services
Linda Patch, Mathematics Consultant, 6–8, NCDPI Instructional Services
Diann Irwin, Section Chief for Behavior Support Services, NCDPI Exceptional Children
Martha Downing, Consultant for the Hearing Impaired, NCDPI, Exceptional Children
Tom Winton, Consultant for the Visual Impaired, NCDPI, Exceptional Children

May 15, 2003, Room 150, Education Building, 2 p.m.–4 p.m.
Attendance
Lou Fabrizio, Director, Division of Accountability Services, NCDPI
Mildred Bazemore, Section Chief for Test Development, NCDPI
Sarah McManus, Lead Consultant for Mathematics Assessments, NCDPI Testing
Laura Kramer, Senior Psychometrician, NCDPI Testing
June Atkinson, Director, Division of Instructional Services, NCDPI
Frances Hoch, Section Chief for Second Languages, ESL, Information
and Computer Skills, NCDPI
Bill Tucci, Section Chief for Mathematics and Science, NCDPI Instructional Services
Toni Meyer, Mathematics Consultant, K–5, NCDPI Instructional Services
Bill Scott, Mathematics Consultant, 9–12, NCDPI Instructional Services
Linda Patch, Mathematics Consultant, 6–8, NCDPI Instructional Services

Test Coordinator's Advisory Committee
May 1, 2003, Room 224, Education Building, 10:45 a.m.–11:15 a.m.
Sarah McManus presented information to the committee about proposed new item formats. Feedback was gathered.

Technical Advisory Committee Meeting
May 8, 2003, SBE lounge, 7th floor, Education Building, 1 p.m.–2 p.m.
Sarah McManus presented information to the committee about proposed new item formats. Feedback was gathered.

Mathematics End-of-Grade—Grade 6 Test Specifications Committee Meeting
August 20, 2003, Room 694, Education Building, 9:00 a.m.–4 p.m.
August 21, 2003, Room 695, Education Building, 9:00 a.m.–11 a.m.
Attendance
Sarah McManus, Lead Consultant for Mathematics Assessments, NCDPI Testing
Toni Meyer, Mathematics Consultant, K–5, NCDPI Instructional Services
Linda Patch, Mathematics Consultant, 6–8, NCDPI Instructional Services
Donna Taylor, Mathematics Consultant, 6–8, NCDPI Instructional Services
Sallie Abbas, Mathematics Consultant, Technical Outreach for Public Schools
Thomas Englehart, Mathematics Consultant, Technical Outreach for Public Schools
Sarah Lawrence, Grade 8 Teacher, Central Middle School, Surry County
Patricia Sickles, Secondary Mathematics Specialist, Durham Public Schools
Karen Vaughan, K–12 Math/Science Curriculum Specialist, Pitt County

NCDPI EOG Mathematics Test Specifications Meeting
September 4, 2003, Room 224, Education Building, 12:30–2:30 p.m.

Attendance (all are members of the NCDPI staff)
Zoe Locklear, Associate Superintendent, Leadership Development and Social Services
Lou Fabrizio, Director, Division of Accountability Services
June Atkinson, Director, Division of Instructional Services
Mary Watson, Director, Division of Accountability Services
Mildred Bazemore, Section Chief for Test Development
Tammy Howard, Section Chief for Testing Policies and Operations
Bill Tucci, Section Chief for Mathematics and Science
Frances Hoch, Section Chief for Second Languages, ESL, Information
and Computer Skills, Instructional Services
David Mills, Section Chief for Areas of Exceptionality, Exceptional Children
Sarah McManus, Lead Consultant for Mathematics Assessments, Testing
Laura Kramer, Senior Psychometrician, Testing
Kelly Burling, Psychometrician, Testing
Pam VanDyk, Technical Writer and Lead Consultant for NCCATS Project, Testing
Toni Meyer, Mathematics Consultant, K–5, Instructional Services
Linda Patch, Mathematics Consultant, 6–8, Instructional Services
Donna Taylor, Mathematics Consultant, 6–8, Instructional Services
Tom Winton, Consultant for the Visual Impaired, Exceptional Children

As needed, this document will be revised. If major changes are needed, the Mathematics End-of-Grade—Grade 6 Test Specifications Committee will meet again.

Draft Test Specifications
The North Carolina End-of-Grade Test of Mathematics—Grade 7
Aligned to the 2003 Mathematics Curriculum

1. Timeline for Aligning End-of-Grade—Grade 7 Tests to the 2003 Mathematics Curriculum [Omitted, as in Grade 3]
2. Purpose and Uses of the Test [Omitted, as in Grade 3]
3. Eligible Students [Omitted, as in Grade 3]

4. Grade 7-Level Competencies in the 2003 Mathematics *Standard Course of Study*

Major Concepts/Skills	Concepts/Skills to Maintain
Computation with rational numbers	Number properties
Ratio and proportion	Percent
Factors and multiples	Transformations in the coordinate plane
Volume and surface area	Probability
Measures of central tendency	
Box plots and histograms	
Equations and inequalities	
Students will solve relevant and authentic problems using appropriate technology and apply these concepts as well as those developed in earlier years	

Strands: Number and Operations, Measurement, Geometry, Data Analysis and Probability, Algebra

COMPETENCY GOAL 1: The learner will understand and compute with rational numbers.

Objectives
1.01 Develop and use ratios, proportions, and percents to solve problems.
1.02 Develop fluency in addition, subtraction, multiplication, and division of rational numbers.

- Analyze computational strategies.
- Describe the effect of operations on size.
- Estimate the results of computations.
- Judge the reasonableness of solutions.

1.03 Develop flexibility in solving problems by selecting strategies and using mental computation, estimation, calculators or computers, and paper and pencil.

COMPETENCY GOAL 2: The learner will understand and use measurement involving two- and three-dimensional figures.

Objectives
2.01 Draw objects to scale and use scale drawings to solve problems.
2.02 Solve problems involving volume and surface area of cylinders, prisms, and composite shapes.

COMPETENCY GOAL 3: The learner will understand and use properties and relationships in geometry.

Objectives
3.01 Using three-dimensional figures:
- Identify, describe, and draw from various views (top, side, front, corner).
- Build from various views.
- Describe cross-sectional views.

3.02 Identify, define, and describe similar and congruent polygons with respect to angle measures, length of sides, and proportionality of sides.
3.03 Use scaling and proportional reasoning to solve problems related to similar and congruent polygons.

COMPETENCY GOAL 4 (Data Analysis and Probability): The learner will understand and use graphs and data analysis.

Objectives
4.01 Collect, organize, analyze, and display data (including box plots and histograms) to solve problems.
4.02 Calculate, use, and interpret the mean, median, mode, range, frequency distribution, and inter-quartile range for a set of data.
4.03 Describe how the mean, median, mode, range, frequency distribution, and inter-quartile range of a set of data affect its graph.
4.04 Identify outliers and determine their effect on the mean, median, mode, and range of a set of data.
4.05 Solve problems involving two or more sets of data using appropriate statistical measures.

COMPETENCY GOAL 5 (Algebra): The learner will demonstrate an understanding of linear relations and fundamental algebraic concepts.

Objectives
5.01 Identify, analyze, and create linear relations, sequences, and functions using symbols, graphs, tables, diagrams, and written descriptions.
5.02 Translate among different representations of algebraic expressions, equations and inequalities.
5.03 Use and evaluate algebraic expressions, linear equations or inequalities to solve problems.
5.04 Develop fluency in the use of formulas to solve problems.

5. Objectives Measured Indirectly
Some objectives are pedagogical in nature and are not intended to be measured on a multiple-choice test. Items will not be written to specifically measure these objectives. However, they are measured indirectly because they need to be mastered before a student is able to fully master other objectives or grade-level competencies.

Some curricular objectives are written in broad terms and are intended to cover a variety of competencies that can be found in other objectives. An item on the Mathematics End-of-Grade—Grade 7 test must be matched to one and only one objective. Therefore, for item-development purposes, items will not be written to specifically measure objectives written in broad terms. However, they are tested indirectly.

Goal and Strand	Objective(s) Measured Indirectly	Reason(s)
1: Number and Operations	1.03	Includes competencies found in other objectives
2: Measurement	none	
3: Geometry	none	
4: Data Analysis & Probability	none	
5: Algebra	none	

6. Miscellaneous Remarks about Objectives

The following table contains remarks made about some of the grade 7 objectives. These remarks are listed to remind members of the Mathematics End-of-Grade—Grade 7 Test Specifications Committee of discussions about the objectives and to assist with item development. The table is not intended to direct instruction. For more information about the objectives, review the *Grade Seven Indicators for the Mathematics* Standard Course of Study available at www.learnnc.org/dpi/instserv.nsf/Category7.

Objective	Remarks
1.01	Includes % increase and % decrease
1.02	Match items to the whole objective
2.02	Cones and pyramids are not included
3.01	Inquire about accessibility for blind and visually impaired students
3.02	Students are not expected to know the triangle congruence/similarity postulates and theorems
4.01	Students are expected to be able to use percents and angle measures when interpreting circle graphs

7. Percent of Items by Goal per Test Form
The test will be designed to meet a percent-by-strand specification rather than a number-per-objective specification.

Goal and Strand	Percent	Priority of Objective Coverage
1: Number & Operation	20–25%	1.01, 1.02
2: Measurement	10–15%	2.01, 2.02
3: Geometry	20–25%	3.03, 3.02, 3.01
4: Data Analysis & Probability	20–25%	4.02, 4.01, 4.05, 4.03, 4.04
5: Algebra	25–30%	5.03, 5.02, 5.01, 5.04

8. Reporting by Strand [Omitted, as in Grade 3]
9. Design of Assessments Aligned to the 2003 Mathematics Curriculum [Omitted, as in Grade 3]
10. Mode of Presentation [Omitted, as in Grade 3]
11. Some Technical Characteristics of Items [Omitted, as in Grade 3]
12. Levels of Difficulty [Omitted, as in Grade 3]
13. Thinking Skills [Omitted, as in Grade 3]

14. Number of Items per Form
Each test form will consist of 80 multiple-choice items. Of these, 49 will be operational items and 31 will be field test items. Operational items will be used to calculate a student's overall test score. Field test items will not be used to calculate a student's test score. Items will be placed in either the Calculator Active or Calculator Inactive part of the test.

	Operational Items	Field Test Items
Calculator Active	36	18
Calculator Inactive	13	13
Total	**49**	**31**

15. Test Format
The test will consist of two parts as shown in the table below.

Calculator Active Part	Calculator Inactive Part
Calculator allowed	No calculator allowed
Approximately 70% of the test	Approximately 30% of the test
Must include multiplication of numbers larger than 2-digit by 2-digit Must include division involving larger than 3-digit numbers divided by 2-digit numbers	Must include 1.01, 1.02
3 sections (18 items each)	2 sections (13 items each)
Graph paper and blank paper will be provided	Graph paper and blank paper will be provided

16. Limited English Proficient Students and Testing Accommodations [Omitted, as in Grade 3]
17. Students with Disabilities and Testing Accommodations [Omitted, as in Grade 3]
18. Accommodations [Omitted, as in Grade 3]
19. Testing Window [Omitted, as in Grade 3]
20. Test Schedule [Omitted, as in Grade 3]
21. Estimated Administration Time [Omitted, as in Grade 6]
22. Additional Assessment Materials [Omitted, as in Grade 3]
23. Calculator Use [Omitted, as in Grade 6]
24. Scoring and Reporting [Omitted, as in Grade 3]
25. Special Requests [Omitted, as in Grade 3]
26. Meetings and Attendance [Omitted, as in Grade 6]

Draft Test Specifications
The North Carolina End-of-Grade Test of Mathematics—Grade 8
Aligned to the 2003 Mathematics Curriculum

1. Timeline for Aligning End-of-Grade—Grade 8 Tests to the 2003 Mathematics Curriculum [Omitted, as in Grade 3]
2. Purpose and Uses of the Test [Omitted, as in Grade 3]
3. Eligible Students [Omitted, as in Grade 3]

4. Grade 8-Level Competencies in the 2003 Mathematics *Standard Course of Study*

Major Concepts/Skills	Concepts/Skills to Maintain
Real numbers	Ratio, proportion, and percent
Linear functions	Factors and multiples
Pythagorean theorem, indirect measurement	Box plots and histograms
Scatterplots	Volume and surface area
Slope	
Equations and inequalities	
Students will solve relevant and authentic problems using appropriate technology and apply these concepts as well as those developed in earlier years	

Strands: Number and Operations, Measurement, Geometry, Data Analysis and Probability, Algebra

COMPETENCY GOAL 1: The learner will understand and compute with real numbers.

Objectives
1.01 Develop number sense for the real numbers.
- Define and use irrational numbers.
- Compare and order.
- Use estimates of irrational numbers in appropriate situations.

1.02 Develop flexibility in solving problems by selecting strategies and using mental computation, estimation, calculators or computers, and paper and pencil.

COMPETENCY GOAL 2: The learner will understand and use measurement concepts.

Objectives
2.01 Determine the effect on perimeter, area or volume when one or more dimensions of two- and three-dimensional figures are changed.
2.02 Apply and use concepts of indirect measurement.

COMPETENCY GOAL 3: The learner will understand and use properties and relationships in geometry.

Objectives
3.01 Represent problem situations with geometric models.

3.02 Apply geometric properties and relationships, including the Pythagorean theorem, to solve problems.
3.03 Identify, predict, and describe dilations in the coordinate plane.

COMPETENCY GOAL 4: The learner will understand and use graphs and data analysis.

Objectives
4.01 Collect, organize, analyze, and display data (including scatterplots) to solve problems.
4.02 Approximate a line of best fit for a given scatterplot; explain the meaning of the line as it relates to the problem and make predictions.
4.03 Identify misuses of statistical and numerical data.

COMPETENCY GOAL 5: The learner will understand and use linear relations and functions.

Objectives
5.01 Develop an understanding of function.

- Translate among verbal, tabular, graphic, and algebraic representations of functions.
- Identify relations and functions as linear or nonlinear.
- Find, identify, and interpret the slope (rate of change) and intercepts of a linear relation.
- Interpret and compare properties of linear functions from tables, graphs, or equations.

5.02 Write an equation of a linear relationship given: two points, the slope and one point on the line, or the slope and y-intercept.
5.03 Solve problems using linear equations and inequalities; justify symbolically and graphically.
5.04 Solve equations using the inverse relationships of addition and subtraction, multiplication and division, squares and square roots, and cubes and cube roots.

5. Objectives Measured Indirectly
Some objectives are pedagogical in nature and are not intended to be measured on a multiple-choice test. Items will not be written to specifically measure these objectives. However, they are measured indirectly because they need to be mastered before a student is able to fully master other objectives or grade level competencies.

Some curricular objectives are written in broad terms and are intended to cover a variety of competencies that can be found in other objectives. An item on the Mathematics End-of-Grade—Grade 8 test must be matched to one and only one objective. Therefore, for item-development purposes, items will not be written to specifically measure objectives written in broad terms. However, they are tested indirectly.

Goal and Strand	Objective(s) Measured Indirectly	Reason(s)
1: Number and Operations	1.02	Includes competencies found in other objectives.
2: Measurement	none	
3: Geometry	none	
4: Data Analysis & Probability	none	
5: Algebra	none	

6. Miscellaneous Remarks about Objectives

The following table contains remarks made about some of the grade 8 objectives. These remarks are listed to remind members of the Mathematics End-of-Grade—Grade 8 Test Specifications Committee of discussions about the objectives and to assist with item development. The table is not intended to direct instruction. For more information about the objectives, one should review the *Grade Eight Indicators for the Mathematics* Standard Course of Study available at www.learnnc.org/dpi/instserv.nsf/Category7.

Objective	Remarks
3.01	Includes geometric probability

7. Percent of Items by Goal per Test Form

The test will be designed to meet a percent-by-strand specification rather than a number-per-objective specification.

Goal and Strand	Percent	Priority of Objective Coverage
1: Number & Operations	10–15%	Equal priority
2: Measurement	10–15%	2.01, 2.02
3: Geometry	10–15%	3.02, 3.01, 3.03
4: Data Analysis & Probability	20–25%	4.01, 4.02, 4.03
5: Algebra	35–40%	5.01, 5.03, 5.04, 5.02

8. Reporting by Strand

After the operational test is administered and standards have been set, Goal Summary Reports will be available for schools. The Goal Summary Report will provide information by Measurement and Geometry (Goals 2 & 3), Data Analysis and Probability (Goal 4), and Algebra and Number and Operations (Goals 1 & 5). Information about a strand/goal can be reported if a minimum of 6 items for that strand/goal are on the test.

9. Design of Assessments Aligned to the 2003 Mathematics Curriculum [Omitted, as in Grade 3]
10. Mode of Presentation [Omitted, as in Grade 3]
11. Some Technical Characteristics of Items [Omitted, as in Grade 3]
12. Levels of Difficulty [Omitted, as in Grade 3]
13. Thinking Skills [Omitted, as in Grade 3]

14. Number of Items per Form
The number of items has not been determined.

15. Test Format
The entire test will be calculator active.

16. Limited English Proficient Students and Testing Accommodations [Omitted, as in Grade 3]
17. Students with Disabilities and Testing Accommodations [Omitted, as in Grade 3]
18. Accommodations [Omitted, as in Grade 3]
19. Testing Window [Omitted, as in Grade 3]
20. Test Schedule [Omitted, as in Grade 3]

21. Estimated Administration Time
Students are expected to complete state tests in the allotted test-administration time (unless a student with a disability or limited English proficiency has the appropriate documentation in accordance with state policies to receive time-modified accommodations).

The administration time has not been determined.

22. Additional Assessment Materials
Each student will be given centimeter graph paper and scratch paper at the beginning of each test administration. A calculator is allowed on the entire test. No formula sheets will be provided. No manipulatives (e.g. rulers, protractors, pattern blocks) will be required.

23. Calculator Use
All students must have access to calculators during the administration of the test. The minimum calculator requirements listed below denote what every calculator must have.

Minimum ("at least") Calculator Requirement
Any four-function calculator with square root function, y^x, π (*pi*), and algebraic logic

Additional features that are not restricted (see below) are **allowed but are not required**.

*****Restrictions*****
Students are not allowed to use calculators with a typewriter-style (QWERTY) keyboard or calculators that include a computer algebraic system (CAS) and are capable of doing symbolic algebra (i.e., factoring, expanding, or simplifying given variable output), *or symbolic calculus. Cell phones, handheld microcomputers, pen-input devices (such as personal digital assistants or tablets), or laptop/notebook computers are not allowed during any of the test administrations*. As curricula and technology change, the policy concerning technology use on North Carolina tests will be reevaluated.

Students are not allowed to share calculators during the test administration.

Students, who regularly use more than one calculator during classroom instructional activities, may be permitted to use more than one calculator during the test administration. Students may use calculators with fraction keys; however, the use of fraction keys without prior training may confuse students and may adversely affect their performance during the test administration.

24. Scoring and Reporting [Omitted, as in Grade 3]
25. Special Requests [Omitted, as in Grade 3]
26. Meetings and Attendance [Omitted, as in Grade 6]

Test Specifications
The North Carolina End-of-Course Test of Mathematics—Algebra I
Aligned to the 2003 Mathematics Curriculum

1. Purpose and Uses

The North Carolina End-of-Course Test—Algebra I (Mathematics End-of-Course—Algebra I) is one component of the North Carolina Testing Program, which includes reading comprehension and mathematics tests in grades three through eight and end-of-course tests in many courses. End-of-Course Tests are required by General Statute 115C-174.10 as a component of the North Carolina Annual Testing Program. 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."*

For school, school system, and state accountability, prediction formulas (first used in 2000–2001) are used to determine growth expectations for end-of-course tests as required by the state-mandated ABCs Accountability Program. The prediction formula is used to determine a student's expected score for each school on each EOC test. Each expected score is determined by students' performance (average scores) on the North Carolina EOG or EOC tests, which serve as predictors of the same students' performance in the EOC course where they are currently enrolled.

2. Timeline for Aligning End-of-Course—Algebra I tests to the 2003 Mathematics Curriculum

Year	Activity
2003–2005	Test Specifications, Item Development
2005–2006	Field Test (paper/pencil and online for use in NCCATS)
2006–2007	1st year of Operational Test (with embedded field test items)

3. Design of Assessments Aligned to the 2003 Mathematics Curriculum

Mathematics End-of-Course—Algebra I will be designed to be accessible to as many students as possible. Item writers will be trained on the universal design principles to ensure that items are accessible to Exceptional Children and Limited English Proficient students. The NCDPI consultants representing the needs of Exceptional Children and Limited English Proficient students will be expected to work with NCDPI Testing and Curriculum staff in reviewing and editing test items. Special consideration will be given to how items will be read aloud to students requiring a "read-aloud" accommodation and for how items will be signed to hearing-impaired students.

Items must be accessible to students taking the test using regular-print test booklets, large-print test booklets, Braille editions, or a computer screen. **Items that cannot be adapted to fit into any one or more of these modes will not be used on the assessment.**

4. Algebra I Competencies in the 2003 Mathematics *Standard Course of Study*

COMPETENCY GOAL 1: The learner will perform operations with numbers and expressions to solve problems.

Objectives

1.01 Write equivalent forms of algebraic expressions to solve problems.
- a) Apply the laws of exponents.
- b) Operate with polynomials.
- c) Factor polynomials.

1.02 Use formulas and algebraic expressions, including iterative and recursive forms, to model and solve problems.

1.03 Model and solve problems using direct variation.

COMPETENCY GOAL 2: The learner will describe geometric figures in the coordinate plane algebraically.

Objectives

2.01 Find the lengths and midpoints of segments to solve problems.

2.02 Use the parallelism or perpendicularity of lines and segments to solve problems.

COMPETENCY GOAL 3: The learner will collect, organize, and interpret data with matrices and linear models to solve problems.

Objectives

3.01 Use matrices to display and interpret data.

3.02 Operate (addition, subtraction, scalar multiplication) with matrices to solve problems.

3.03 Create linear models for sets of data to solve problems.
- a) Interpret constants and coefficients in the context of the data.
- b) Check the model for goodness-of-fit and use the model, where appropriate, to draw conclusions or make predictions.

COMPETENCY GOAL 4: The learner will use relations and functions to solve problems.

Objectives

4.01 Use linear functions or inequalities to model and solve problems; justify results.
- a) Solve using tables, graphs, and algebraic properties.
- b) Interpret constants and coefficients in the context of the problem.

4.02 Graph, factor, and evaluate quadratic functions to solve problems.

4.03 Use systems of linear equations or inequalities in two variables to model and solve problems. Solve using tables, graphs, and algebraic properties; justify results.

4.04 Graph and evaluate exponential functions to solve problems.

5. Item Formats

All items will be in multiple-choice format with only one correct answer choice and three plausible distractors. Each item will be matched to only one objective.

6. Number of Items per Form

The exact number of items on the test **will not** be made final until the field test results have been studied. The operational test will be designed so that the total number of items will be about 80. For example, the operational test may have 60 operational items and 20 field test items.

7. Objectives Measured Indirectly

Some objectives are pedagogical in nature and are not intended to be measured on a multiple-choice test. Items will not be written to specifically measure these objectives. However, they are measured indirectly because they need to be mastered before a student is able to fully master other objectives or grade-level competencies.

Some curricular objectives are written in broad terms and are intended to cover a variety of competencies that can be found in other objectives. An item on the Mathematics End-of-Course—Algebra I test must be matched to one and only one objective. Therefore, for item-development purposes, items will not be written to specifically measure objectives written in broad terms. However, they are tested indirectly.

Goal and Strand	Objective(s) Measured Indirectly	Reason(s)
1: Number and Operations	N/A	
2: Geometry and Measurement	N/A	
3: Data Analysis & Probability	N/A	
4: Algebra	N/A	

8. Miscellaneous Remarks about Objectives

The following table contains remarks made about some of the Algebra I objectives. These remarks are listed to remind members of the Mathematics End-of-Course—Algebra I Test Specifications Committee of discussions about the objectives and to assist with item development. The table is not intended to direct instruction. For more information about the objectives review the *Algebra I Indicators for the Mathematics* Standard Course of Study available at www.learnnc.org/dpi/instserv.nsf/category7.

Objective	Remarks
1.01b	Limit to division by monomials; simplifying radicals is not necessary because decimal approximations should be used; students are not expected to be able to work with rational exponents.
4.03	Students are expected to be able to justify results in an algebraic proof
4.04	Exponential functions involve growth and decay

9. Level of Difficulty
Prior to field testing, North Carolina educators will use professional judgments to place an item into one of three categories: Easy, Medium, or Hard.

Level of Difficulty	Percent of Items per Form
Easy	25%
Medium	50%
Hard	25%

The average p-value (the percent of students getting the item correct) of all the items on the test should preferably be in the vicinity of .625. However, as a minimum, the overall p-value will be the average of the operational item pool.

10. Thinking Skills
The thinking-skills framework used in developing Mathematics End-of-Course—Algebra I is from *Dimensions of Thinking* by Robert J. Marzano and others (1988). Items will be categorized as requiring the use of lower-order or higher-order thinking skills.

Lower-Order Thinking Skills	Higher-Order Thinking Skills
Knowledge	Analyzing
Organizing	Generating
Applying	Integrating
	Evaluating
40% of test	60% of test

11. Test Format
The test will consist of one part and a calculator can be allowed. Students will be provided with graph paper.

12. Percent of Items by Goal per Test Form
The test will be designed to meet a percent-by-strand specification rather than a number-per-objective specification.

Goal and Strand	Percent	Priority of Objective Coverage
1: Number & Operation	20–25 %	1.01, 1.02, 1.03
2: Measurement & Geometry	10–15 %	2.02, 2.01
3: Data Analysis & Probability	30–35 %	3.02, 3.01, 3.03 (equal weighting)
4: Algebra	35–40 %	4.01, 4.03, 4.02, 4.04

13. Reporting by Strand
After the operational test is administered and standards have been set, Goal Summary Reports will be available for schools. The Goal Summary Report will provide information by strand.

14. Calculator Use
All students must have access to calculators during the administration of the test. The minimum calculator requirements listed below denote what every calculator must have.

Minimum ("at least") Calculator Requirement
Graphing Calculator

Additional features that are not restricted (see below) are **allowed but are not required**.

*****Restrictions*****
Students are not allowed to use calculators with a typewriter-style (QWERTY) keyboard or calculators that include a computer algebraic system (CAS) and are capable of doing symbolic algebra (i.e., factoring, expanding, or simplifying given variable output), or symbolic calculus. *Cell phones, handheld microcomputers, pen-input devices (such as personal digital assistants or tablets), or laptop/notebook computers are not allowed during any of the test administrations.* As curricula and technology change, the policy concerning technology use on North Carolina tests will be reevaluated.

Students are not allowed to share calculators during the test administration.

Students, who regularly use more than one calculator during classroom instructional activities, may be permitted to use more than one calculator during the test administration. Students may use calculators with fraction keys; however, the use of fraction keys without prior training may confuse students and may adversely affect their performance during the test administration.

15. Additional Assessment Materials
Each student will be given graph paper and scratch paper at the start of each test administration. A calculator is allowed on the test. No formula sheets will be required.

16. Meetings and Attendance
Preliminary meetings were held to discuss how to align tests to the 2003 curriculum. Decisions made at these meetings helped to guide discussions at the Mathematics End-of-Course—Algebra I Test Specifications meeting held on October 14 and 15, 2003.

Preliminary Meetings to Discuss Aligning Tests to the 2003 Math Curriculum
April 11, 2003, room 694, Education Building, 9 a.m.–11 a.m.
Attendance
Sarah McManus, Lead Consultant for Mathematics Assessments, NCDPI Testing
Laura Kramer, Senior Psychometrician, NCDPI Testing
Bill Tucci, Section Chief for Mathematics and Science, NCDPI Instructional Services
Toni Meyer, Mathematics Consultant, K–5, NCDPI Instructional Services
Bill Scott, Mathematics Consultant, 9–12, NCDPI Instructional Services
Linda Patch, Mathematics Consultant, 6–8, NCDPI Instructional Services

April 30, 2003, 5th floor library, Education Building, 11 a.m.–2:30 p.m.
Attendance

Sarah McManus, Lead Consultant for Mathematics Assessments, NCDPI Testing
Kevin Murphy, Operations Manager, NCDPI Testing
Toni Meyer, Mathematics Consultant, K–5, NCDPI Instructional Services
Bill Scott, Mathematics Consultant, 9–12, NCDPI Instructional Services
Linda Patch, Mathematics Consultant, 6–8, NCDPI Instructional Services
Diann Irwin, Section Chief for Behavior Support Services, NCDPI Exceptional Children
Martha Downing, Consultant for the Hearing Impaired, NCDPI, Exceptional Children
Tom Winton, Consultant for the Visual Impaired, NCDPI, Exceptional Children

May 15, 2003, Room 150, Education Building, 2 p.m.–4 p.m.
Attendance
Lou Fabrizio, Director, Division of Accountability Services, NCDPI
Mildred Bazemore, Section Chief for Test Development, NCDPI
Sarah McManus, Lead Consultant for Mathematics Assessments, NCDPI Testing
Laura Kramer, Senior Psychometrician, NCDPI Testing
June Atkinson, Director, Division of Instructional Services, NCDPI
Frances Hoch, Section Chief for Second Languages, ESL, Information
and Computer Skills, NCDPI
Bill Tucci, Section Chief for Mathematics and Science, NCDPI Instructional Services
Toni Meyer, Mathematics Consultant, K–5, NCDPI Instructional Services
Bill Scott, Mathematics Consultant, 9–12, NCDPI Instructional Services
Linda Patch, Mathematics Consultant, 6–8, NCDPI Instructional Services

Test Coordinator's Advisory Committee
May 1, 2003, Room 224, Education Building, 10:45 a.m.–11:15 a.m.
Sarah McManus presented information to the committee about proposed new item formats.
Feedback was gathered.

Technical Advisory Committee Meeting
May 8, 2003, SBE lounge, 7th floor, Education Building, 1 p.m.–2 p.m.
Sarah McManus presented information to the committee about proposed new item formats.
Feedback was gathered.

Mathematics End-of-Course—Algebra I Test Specifications Committee Meeting
October 14, 2003, Room 228 Education Building 9:00 a.m.–3:30 p.m.
October 15, 2003, Room 228, Education Building 9:00 a.m.–2 p.m.
Attendance
Sarah McManus, Lead Consultant for Mathematics Assessments, NCDPI Testing
Bill Scott, Mathematics Consultant, 9–12, NCDPI Instructional Services
Linda Patch, Mathematics Consultant, 6–8, NCDPI Instructional Services
Robert Brown, Test Development Consultant, Technical Outreach for Public Schools
Thomas Englehart, Mathematics Consultant, Technical Outreach for Public Schools
Sallie Abbas, Mathematics Consultant, Technical Outreach for Public Schools
Vivien Windley, Mathematics Teacher, Southern Vance High School
Cheryl Hassell, Secondary Education Curriculum Specialist, Beaufort County Schools

Test Specifications

The North Carolina End-of-Course—Geometry
Aligned to the 2003 Mathematics Curriculum

1. Purpose and Uses [Omitted – Same as Algebra I]

2. Timeline for Aligning End-of-Course Test—Geometry to the 2003 Mathematics Curriculum [Omitted – Same as Algebra I]

3. Design of Assessments Aligned to the 2003 Mathematics Curriculum [Omitted – Same as Algebra I]

4. End-of-Course—Geometry Competencies in the 2003 Mathematics *Standard Course of Study*

COMPETENCY GOAL 1: The learner will perform operations with real numbers to solve problems. *Objectives:*

1.01 Use the trigonometric ratios to model and solve problems involving right triangles.
1.02 Use length, area, and volume of geometric figures to solve problems. Include arc length, area of sectors of circles; lateral area, surface area, and volume of three-dimensional figures; and perimeter, area, and volume of composite figures.
1.03 Use length, area, and volume to model and solve problems involving probability.

COMPETENCY GOAL 2: The learner will use geometric and algebraic properties of figures to solve problems and write proofs. *Objectives:*

2.01 Use logic and deductive reasoning to draw conclusions and solve problems.
2.02 Apply properties, definitions, and theorems of angles and lines to solve problems and write proofs.
2.03 Apply properties, definitions, and theorems of two-dimensional figures to solve problems and write proofs:
- a) Triangles.
- b) Quadrilaterals.
- c) Other polygons.
- d) Circles.

2.04 Develop and apply properties of solids to solve problems.

COMPETENCY GOAL 3: The learner will transform geometric figures in the coordinate plane algebraically. *Objectives:*

3.01 Describe the transformation (translation, reflection, rotation, dilation) of polygons in the coordinate plane in simple algebraic terms.
3.02 Use matrix operations (addition, subtraction, multiplication, scalar multiplication) to describe the transformation of polygons in the coordinate plane.

5. Item Formats [Omitted – Same as Algebra I]

6. Number of Items per Form [Omitted – Same as Algebra I]

7. Objectives Measured Indirectly [Omitted – Same as Algebra I]

8. Miscellaneous Remarks about Objectives

The following table contains remarks made about some of the End-of-Course Test—Geometry objectives. These remarks are listed to remind members of the North Carolina End-of-Course Test—Geometry Test Specifications Committee of discussions about the objectives and to assist with item development. The table is not intended to direct instruction. For more information about the objectives review the *Geometry Indicators for the Mathematics* Standard Course of Study available at www.learnnc.org/dpi/instserv.nsf/category7.

Objective	Remarks
1.02	Distance can involve 3-D or 2-D figures; Use 2-D or 3-D composite figures
2.03	Involves theorems related to congruence and similarity
3.02	Multiplication involves variations of the identity matrix; Rotations involve multiples of 90 degrees

9. Level of Difficulty [Omitted – Same as Algebra I]

10. Thinking Skills

The thinking skills framework used in developing North Carolina High School Comprehensive Test—Mathematics is from *Dimensions of Thinking* by Robert J. Marzano and others (1988). Items will be categorized as requiring the use of lower-order or higher-order thinking skills.

Lower-Order Thinking Skills	Higher-Order Thinking Skills
Knowledge	Analyzing
Organizing	Generating
Applying	Integrating
	Evaluating
35% of test	**65% of test**

11. Test Format [Omitted – Same as Algebra I]

12. Percent of Items by Goal per Test Form

The test will be designed to meet a percent-by-strand specification rather than a number-per-objective specification.

Goal and Strand	Percent	Priority of Objective Coverage
1: Number & Operation	25–30 %	1.02, 1.01, 1.03
2: Geometry	50–55 %	2.03, 2.02, 2.04, 2.01,
3: Algebra	20–25 %	3.01, 3.02

13. Reporting by Strand [Omitted – Same as Algebra I]

14. Calculator Use

All students must have access to calculators during the administration of the test. The minimum calculator requirements listed below denote what every calculator must have.

Minimum ("at least") Calculator Requirement
Scientific Calculator

Additional features that are not restricted (see below) are **allowed but are not required**.

*****Restrictions*****
Students are <u>*not*</u> *allowed to use calculators with a typewriter-style (QWERTY) keyboard or calculators that include a computer algebraic system (CAS) and are capable of doing symbolic algebra* (*i.e.,* factoring, expanding, or simplifying given variable output), or symbolic calculus. *Cell phones, handheld microcomputers, pen-input devices (such as personal digital assistants or tablets), or laptop/notebook computers are* <u>*not*</u> *allowed during any of the test administrations.* As curricula and technology change, the policy concerning technology use on North Carolina tests will be reevaluated.

Students are <u>not</u> allowed to share calculators during the test administration.

Students, who regularly use more than one calculator during classroom instructional activities, <u>may</u> be permitted to use more than one calculator during the test administration. Students <u>may</u> use calculators with fraction keys; however, the use of fraction keys without prior training may confuse students and may adversely affect their performance during the test administration.

15. Additional Assessment Materials [Omitted – Same as Algebra I]

16. Meetings and Attendance [Omitted – Same as Algebra I]

Test Specifications
The North Carolina End-of-Course Test of Mathematics—Algebra II
Aligned to the 2003 Mathematics Curriculum

1. Purpose and Uses [Omitted – Same as Algebra I]

2. Timeline for Aligning End-of-Course—Algebra II Tests to the 2003 Mathematics Curriculum [Omitted – Same as Algebra I]

3. Design of Assessments Aligned to the 2003 Mathematics Curriculum [Omitted – Same as Algebra I]

4. Algebra II Competencies in the 2003 Mathematics *Standard Course of Study*

COMPETENCY GOAL 1: The learner will perform operations with complex numbers, matrices, and polynomials.

Objectives

1.01 Simplify and perform operations with rational exponents and logarithms (common and natural) to solve problems.
1.02 Define and compute with complex numbers.
1.03 Operate with algebraic expressions (polynomial, rational, complex fractions) to solve problems.
1.04 Operate with matrices to model and solve problems.
1.05 Model and solve problems using direct, inverse, combined and joint variation.

COMPETENCY GOAL 2: The learner will use relations and functions to solve problems.

Objectives

2.01 Use the composition and inverse of functions to model and solve problems; justify results.
2.02 Use quadratic functions and inequalities to model and solve problems; justify results.
   a) Solve using tables, graphs, and algebraic properties.
   b) Interpret the constants and coefficients in the context of the problem.
2.03 Use exponential functions to model and solve problems; justify results.
   a) Solve using tables, graphs, and algebraic properties.
   b) Interpret the constants, coefficients, and bases in the context of the problem.
2.04 Create and use best-fit mathematical models of linear, exponential, and quadratic functions to solve problems involving sets of data.
   a) Interpret the constants, coefficients, and bases in the context of the data.
   b) Check the model for goodness-of-fit and use the model, where appropriate, to draw conclusions or make predictions.
2.05 Use rational equations to model and solve problems; justify results.
   a) Solve using tables, graphs, and algebraic properties.
   b) Interpret the constants and coefficients in the context of the problem.
   c) Identify the asymptotes and intercepts graphically and algebraically.

2.06 Use cubic equations to model and solve problems.
  a) Solve using tables and graphs.
  b) Interpret constants and coefficients in the context of the problem.

2.07 Use equations with radical expressions to model and solve problems; justify results.
  a) Solve using tables, graphs, and algebraic properties.
  b) Interpret the degree, constants, and coefficients in the context of the problem.

2.08 Use equations and inequalities with absolute value to model and solve problems; justify results.
  a) Solve using tables, graphs, and algebraic properties.
  b) Interpret the constants and coefficients in the context of the problem.

2.09 Use the equations of parabolas and circles to model and solve problems; justify results.
  a) Solve using tables, graphs, and algebraic properties.
  b) Interpret the constants and coefficients in the context of the problem.

2.10 Use systems of two or more equations or inequalities to model and solve problems; justify results. Solve using tables, graphs, matrix operations, and algebraic properties.

5. Item Formats [Omitted – Same as Algebra I]

6. Number of Items per Form [Omitted – Same as Algebra I]

7. Objectives Measured Indirectly [Omitted – Same as Algebra I]

8. Miscellaneous Remarks about Objectives

The following table contains remarks made about some of the Algebra II objectives. These remarks are listed to remind members of the Mathematics End-of-Course—Algebra II Test Specifications Committee of discussions about the objectives and to assist with item development. The table is not intended to direct instruction. For more information about the objectives review the *Algebra II Indicators for the Mathematics* Standard Course of Study available at www.learnnc.org/dpi/instserv.nsf/category7.

Objective	**Remarks**
2.06	Polynomials up through cubic
2.10	No explicit linear programming

9. Level of Difficulty [Omitted – Same as Algebra I]

10. Thinking Skills [Omitted – Same as Algebra I]

11. Test Format [Omitted – Same as Algebra I]

12. Percent of Items by Goal per Test Form

The test will be designed to meet a percent-by-strand specification rather than a number-per-objective specification.

Goal and Strand	Percent	Priority of Objective Coverage
1: Number & Operations	25–30 %	1.03, 1.01, 1.04, 1.05, 1.02
2: Algebra	70–75 %	2.03, 2.02, 2.10, 2.06, 2.09, 2.05, 2.01, 2.04, 2.07, 2.08

13. Reporting by Strand [Omitted – Same as Algebra I]

14. Calculator Use [Omitted – Same as Algebra I]

15. Additional Assessment Materials [Omitted – Same as Algebra I]

16. Meetings and Attendance

Preliminary meetings were held to discuss how to align tests to the 2003 curriculum. Decisions made at these meetings helped to guide discussions at the Mathematics End-of-Course—Algebra II Test Specifications meeting held on October 14 and 15, 2003.

Preliminary Meetings to Discuss Aligning tests to the 2003 Math Curriculum

April 11, 2003, room 694, Education Building, 9 a.m.–11 a.m.

Attendance

Sarah McManus, Lead Consultant for Mathematics Assessments, NCDPI Testing
Laura Kramer, Senior Psychometrician, NCDPI Testing
Bill Tucci, Section Chief for Mathematics and Science, NCDPI Instructional Services
Toni Meyer, Mathematics Consultant, K–5, NCDPI Instructional Services
Bill Scott, Mathematics Consultant, 9–12, NCDPI Instructional Services
Linda Patch, Mathematics Consultant, 6–8, NCDPI Instructional Services

April 30, 2003, 5th floor library, Education Building, 11 a.m.–12:30 p.m.

Attendance

Sarah McManus, Lead Consultant for Mathematics Assessments, NCDPI Testing
Kevin Murphy, Operations Manager, NCDPI Testing
Toni Meyer, Mathematics Consultant, K–5, NCDPI Instructional Services
Bill Scott, Mathematics Consultant, 9–12, NCDPI Instructional Services
Linda Patch, Mathematics Consultant, 6–8, NCDPI Instructional Services
Diann Irwin, Section Chief for Behavior Support Services, NCDPI Exceptional Children
Martha Downing, Consultant for the Hearing Impaired, NCDPI, Exceptional Children
Tom Winton, Consultant for the Visual Impaired, NCDPI, Exceptional Children

May 15, 2003, Room 150, Education Building, 2 p.m.–4 p.m.

Attendance

Lou Fabrizio, Director, Division of Accountability Services, NCDPI
Mildred Bazemore, Section Chief for Test Development, NCDPI
Sarah McManus, Lead Consultant for Mathematics Assessments, NCDPI Testing
Laura Kramer, Senior Psychometrician, NCDPI Testing

June Atkinson, Director, Division of Instructional Services, NCDPI
Frances Hoch, Section Chief for Second Languages, ESL, Information and Computer Skills, NCDPI
Bill Tucci, Section Chief for Mathematics and Science, NCDPI Instructional Services
Toni Meyer, Mathematics Consultant, K–5, NCDPI Instructional Services
Bill Scott, Mathematics Consultant, 9–12, NCDPI Instructional Services
Linda Patch, Mathematics Consultant, 6–8, NCDPI Instructional Services

Test Coordinator's Advisory Committee
May 1, 2003, Room 224, Education Building, 10:45 a.m.–11:15 a.m.
Sarah McManus presented information to the committee about proposed new item formats. Feedback was gathered.

Technical Advisory Committee Meeting
May 8, 2003, SBE lounge, 7th floor, Education Building, 1 p.m.–2 p.m.
Sarah McManus presented information to the committee about proposed new item formats. Feedback was gathered.

Mathematics End-of-Course—Algebra II Test Specifications Committee Meeting
October 14, 2003, Room 228 Education Building 9:00 a.m.–3:30 p.m.
October 15, 2003, Room 228, Education Building 9:00 a.m.–2 p.m.
<u>Attendance</u>
Sarah McManus, Lead Consultant for Mathematics Assessments, NCDPI Testing
Bill Scott, Mathematics Consultant, 9–12, NCDPI Instructional Services
Linda Patch, Mathematics Consultant, 6–8, NCDPI Instructional Services
Robert Brown, Test Development Consultant, Technical Outreach for Public Schools
Thomas Englehart, Mathematics Consultant, Technical Outreach for Public Schools
Sallie Abbas, Mathematics Consultant, Technical Outreach for Public Schools
Vivien Windley, Mathematics teacher Southern Vance High School
Cheryl Hassell, Secondary Education Curriculum Specialist, Beaufort County Schools

Mathematics End-of-Course – Algebra II Test Specifications Committee Meeting
March 10, 2005, Room 228, Education Building, 10:00 a.m.–12:00 p.m.
<u>Attendance</u>
Michael Gallagher, Lead Consultant for Mathematics Assessments, NCDPI Testing
Bill Scott, Mathematics Consultant, 9–12, NCDPI Instructional Services
Robert Brown, Test Development Consultant, Technical Outreach for Public Schools
Thomas Englehart, Mathematics Consultant, Technical Outreach for Public Schools
Vivien Windley, Mathematics Curriculum Specialist, Vance County Schools
Cheryl Hassell, Secondary Education Curriculum Specialist, Beaufort County Schools

Appendix C – Creation of the Developmental Scale

The Developmental Scale for the North Carolina End-of-Grade Mathematics Tests, Third Edition

Alan Nicewander, Ph.D.
Howard Mitzel, Ph.D.
Susan Lottridge, Ph.D.
Don Murray, B.S.

Submitted to the
North Carolina Department of Education

December 13, 2006

Pacific Metrics Corporation
585 Cannery Row, Suite 201
Monterey, California 93940

The Developmental Scale for the North Carolina End-of-Grade Mathematics Tests, Third Edition

This technical report describes the results and methods used by Pacific Metrics Corporation to derive the developmental scale for the North Carolina End-of-Grade Mathematics Tests, Third Edition. The results and methods of equating the Third Edition scale scores to the Second Edition scale scores are also described. Pacific Metrics used the methods already in place by North Carolina to create the vertical scale, as described in the *The North Carolina Mathematics Tests Technical Document, Chapter 4* (Bazemore, Kramer, Yelton, & Brown, 2006) and *Appendix C* (Thissen, Sathy, Edwards, & Flora, 2006) to that report. The article by Williams, Pommerich, and Thissen (1998) was also used as a reference.

Third Edition Developmental Scale

Table 1 presents the Third Edition developmental scale for the population. Grade 5 was the base grade for the development scale, using a mean of 350 and standard deviation of 10. To create the developmental scale, the same items (called linking sets) were administered to students in adjacent grades. The linking set in the lower grade was operational (i.e., items contributed to student scores) for that grade, while the linking set in the upper grade was not operational for that grade. The linking sets consisted of 10, 14, or 18 items.

As shown in Table 1, the mean scores increased rapidly between grades 3P and 5, and then increased more slowly for the remaining grades. The smallest growth occurred between grade 5 and grade 6, and the largest growth occurred between grades 3P and 3. The mathematics underlying the developmental scale creation appear at the end of this report.

Table 1. Developmental scale means and standard deviations derived from Spring 2006 item calibration for the North Carolina End-of-Grade Tests of Mathematics, Third Edition

	Population	
Grade	**Mean**	**Standard Deviation**
3 Pretest	326.98	12.69
3	339.44	10.97
4	345.26	10.24
5	350.00	10.00
6	351.45	10.41
7	353.66	10.15
8	355.42	9.99

The values for the developmental scale are based upon Item Response Theory (IRT) estimates of differences between adjacent-grade mean thetas (θ) and ratios of adjacent-grade standard deviations of θ. The three parameter logistical model was used to estimate item and person parameters. BILOG-MG software version 3.0 (Zimowski, Muraki, Mislevy, & Bock, 2002) was used. In BILOG-MG, the below grade was considered the reference group and thus its population mean and standard deviation were set to 0 and 1, respectively. The values of the above grade mean and standard deviation were estimated using the scored data and the IRT parameter estimates. These parameters were provided in the BILOG-MG output and did not require independent calculation.

Individual runs in BILOG-MG were conducted for each of the grade pair forms. Grades 3P through 7 had six test forms and grade 8 had four test forms, resulting in a total of 40 adjacent-grade forms. Under the assumption of equivalent groups, the form results were averaged within grade pairs to produce one set of values per adjacent grade. Table 2 shows the average difference in adjacent-grade means and standard deviation ratios. The value from one grade 5-6 pair was dropped as an outlier because the mean value was less than half of the other mean values for that set of grade-pairs. Table 3 shows the mean difference and standard deviation ratio for each adjacent-grade form.

Table 2. Average mean difference in standard deviation units of the lower grade and average standard deviation ratios derived from the Spring 2006 item calibrations for the North Carolina EOG Tests of Mathematics, Third Edition

Grades	Average Mean Difference	Average Standard Deviation Ratio	Number of Grade-Pair Forms
Pre 3–3	0.982	0.864	6
3–4	0.531	0.933	6
4–5	0.463	0.977	6
5–6*	0.145	1.041	5
6–7	0.212	0.975	6
7–8	0.174	0.984	4

* For grades 5–6 the lowest mean difference (.05) was dropped from the average. Including the mean difference changes the mean ratio to 0.129 and the standard deviation ratio to 1.05.

Table 3. Values for adjacent-grade means in standard deviation units of the lower grade and standard deviation ratios, derived from the Spring 2006 item calibrations for the North Carolina EOG Tests of Mathematics, Third Edition

Grades and Forms	Mean (μ) Difference	(σ) Ratio
Pre3K–3K	0.958	0.795
Pre3K–3L	0.943	0.883
Pre3K–3P	1.124	0.888
Pre3L–3M	1.015	0.895
Pre3L–3N	0.766	0.784
Pre3L–3O	1.083	0.942
3M–4L	0.517	0.922
3M–4M	0.567	0.899
3M–4O	0.527	0.944
3N–4K	0.432	0.923
3N–4N	0.503	0.907
3N–4P	0.642	1.002
4L–5K	0.480	0.981
4L–5N	0.466	0.973
4L–5P	0.438	0.989
4M–5L	0.414	0.990
4M–5M	0.434	0.910
4M–5O	0.548	1.021
5K–6L	0.155	1.025
5K–6N	0.156	1.000
5K–6P*	0.050	1.096
5M–6O	0.116	1.104
5O–6K	0.196	1.042
5O–6M	0.101	1.036
6O–7K	0.132	0.924
6O–7N	0.240	0.917
6O–7O	0.253	0.991
6P–7L	0.206	1.053
6P–7M	0.195	0.964
6P–7P	0.243	1.001
7L–8L	0.185	0.978
7L–8M	0.159	0.967
7N–8K	0.198	1.015
7N–8N	0.155	0.975

*Dropped, as an outlier, from further analyses

Comparison of Third Edition Developmental Scale to First and Second Edition Scales

Table 4 presents the mean scale scores by grade for the First, Second, and Third Edition. More average growth occurred from grade 3P to grade 8 in the First and Second Editions (43.7 and 32.74, respectively) than in the Third Edition (28.44). The growth per grade for each of the First and Second Edition scales tended to be more constant than the Third Edition. In the Third Edition, the most growth occurred in grades 2 through 4 and much less growth occurred in grades 5 through 8.

Table 4. Comparison of the population means and standard deviations for the First, Second, and Third Editions of the North Carolina End-of-Grade Tests of Mathematics

	First Edition (1992)		Second Edition (2000)		Third Edition (2006)	
Grade	**Mean**	**Standard Deviation**	**Mean**	**Standard Deviation**	**Mean**	**Standard Deviation**
Pre 3	131.6	7.8	234.35	9.66	326.98	12.69
3	143.5	11.1	248.27	9.86	339.44	10.97
4	152.9	10.1	252.90	10.65	345.26	10.24
5	159.5	10.1	255.99	12.78	350.00	10.00
6	165.1	11.2	259.95	11.75	351.45	10.41
7	171.0	11.5	263.36	12.46	353.66	10.15
8	175.3	11.9	267.09	12.83	355.42	9.99

To facilitate comparison of the growth between grades among the First, Second, and Third Edition, Figure 1 presents the mean scores plotted together. To place the First, Second, and Third Edition scores on similar scales, a value of 200 was added to the First Edition scores and a value of 100 was added to the Second Edition scores. As shown in Figure 1, the growth of the Third Edition mean scores is similar to the growth of the Second Edition mean scores, and quite different from the First Edition mean scores.

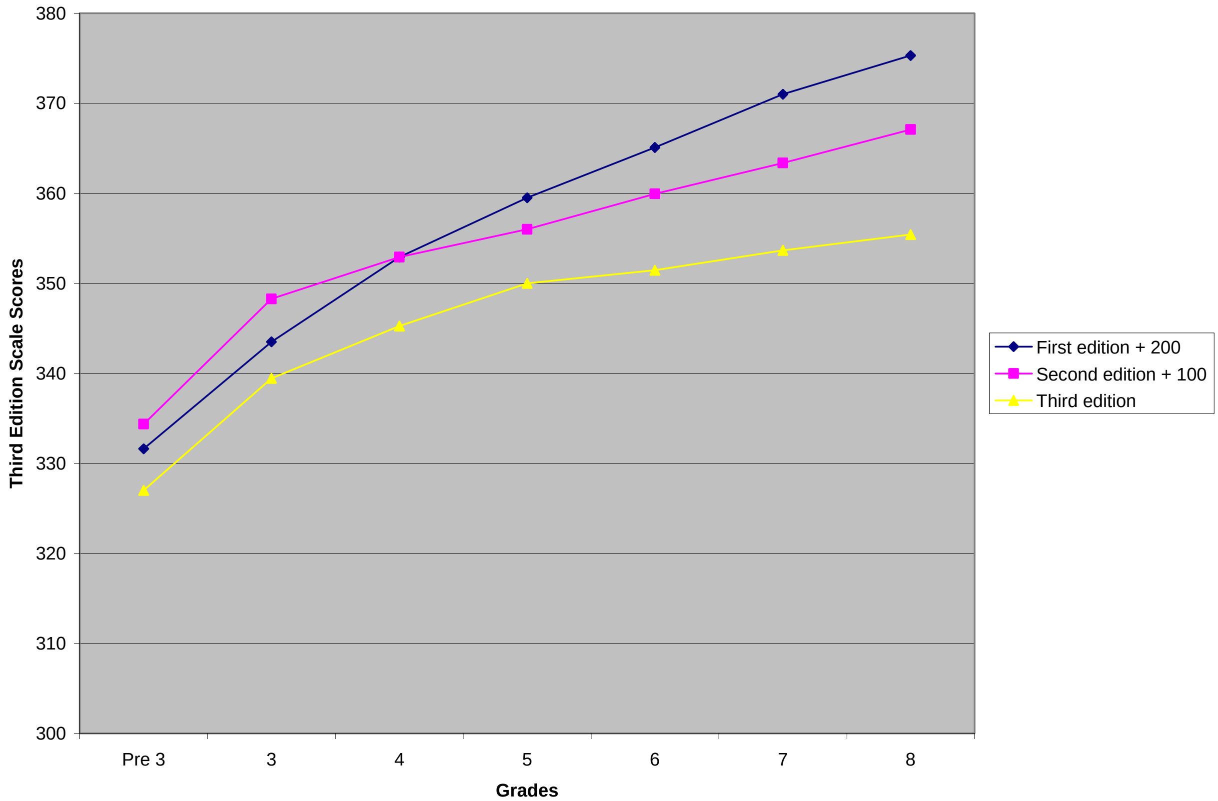


Figure 1. Comparison of growth curves between the First, Second, and Third Editions of the North Carolina End-of-Grade Tests of Mathematics.

Equating the Scales for the Second and Third Editions of the North Carolina End-of-Grade Tests of Mathematics

The Third Edition scales were statistically moderated to the Second Edition scales using equipercentile equating. While the term "equating" is used throughout this section, the term use is technically inadequate as it should only be applied to tests with parallel content (Mislevy, 1992). The Third Edition and Second Edition tests presumably assess slightly different content areas because they represent curriculum updates by the state. However, the equipercentile method is an equating process, and therefore it is called as such throughout this document. The equating process was conducted at each grade and on data gathered in field tests.

As indicated in Table 5, two Second Edition forms and either three or eight Third Edition forms were used in the equating process. The two Second Edition forms consisted of the same items, but with different ordering.

Table 5. Number of field test forms for NC EOG Mathematics Tests Grades 3P-8

	Third Edition Forms		Second Edition Forms	
Grade Level	**#**	**Names**	**#**	**Names**
Pre 3	8	1–8	2	9–10
3	8	1–8	2	9–10
4	8	1–8	2	9–10
5	8	1–8	2	9–10
6	8	1–8	2	9–10
7	3	1–3	2	4–5
8	3	1–3	2	4–5

Data from the forms were combined using the following procedure. Within a grade, examinees who took either of the Second Edition forms received a raw total score which was then converted to the Second Edition scale score. A cumulative frequency distribution was calculated on the scale score data for that grade. Examinees who took a Third Edition form received a raw total score, which was then converted to the Third Edition scale score for that form and grade. The scale data across forms in a grade were then combined, and a cumulative frequency distribution was computed. Equipercentile equating was then conducted on the cumulative frequency distributions of the Second Edition and Third Edition scale scores for each grade using procedures described in Kolen and Brennan (1995, p. 44). Figure 2 and Table 6 present the results of the equipercentile equating.

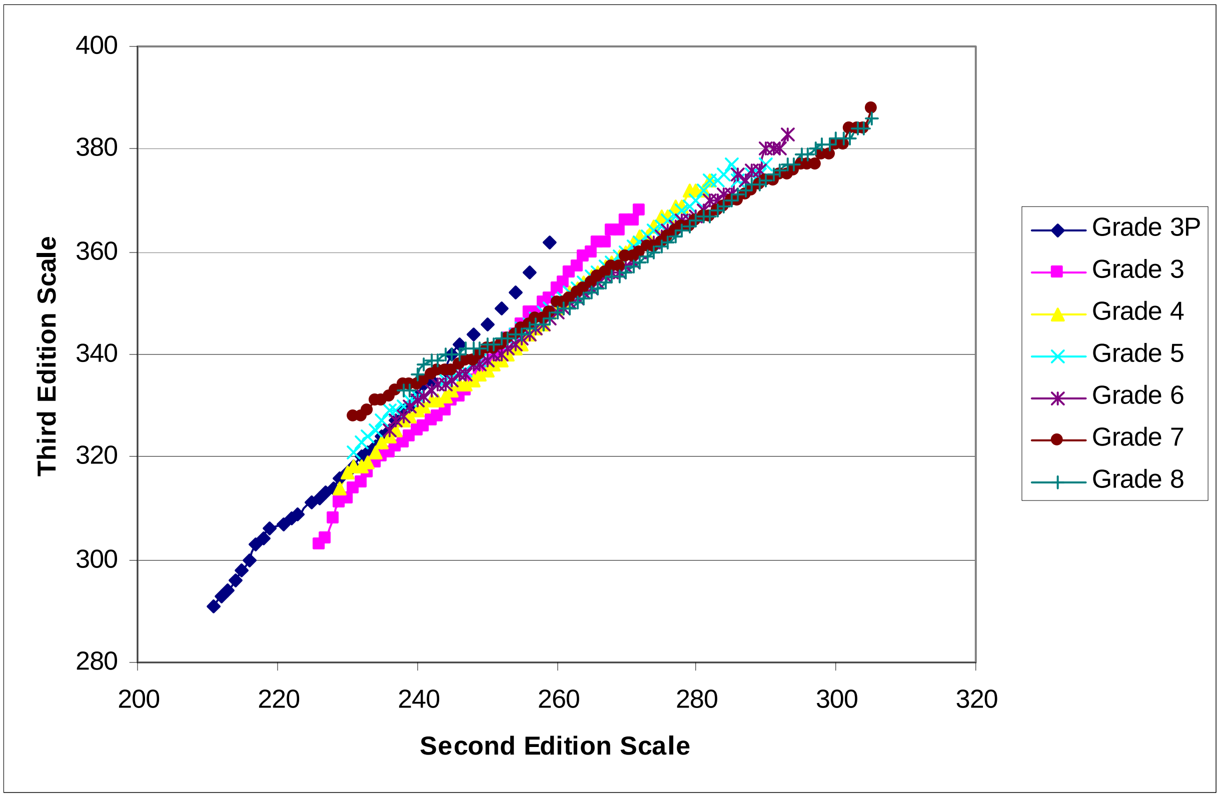


Figure 2. Equipercentile equating functions between the Second and Third Editions of the North Carolina EOG Tests of Mathematics scales for Grades 3P–8.

Table 6. Translation table for "equating" Third Edition scale scores to the Second Edition scale scores of the North Carolina EOG Tests of Mathematics for Grade 3P–8

	Third Edition						
Second Edition	Grade 3P	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
211	291	.	.	.	.	.	.
212	293	.	.	.	.	.	.
213	294	.	.	.	.	.	.
214	296	.	.	.	.	.	.
215	298	.	.	.	.	.	.
216	300	.	.	.	.	.	.
217	303	.	.	.	.	.	.
218	304	.	.	.	.	.	.
219	306	.	.	.	.	.	.
220	307	.	.	.	.	.	.
221	307	.	.	.	.	.	.
222	308	.	.	.	.	.	.
223	309	.	.	.	.	.	.
224	310	.	.	.	.	.	.
225	311	.	.	.	.	.	.
226	312	303	.	.	.	.	.
227	313	304	.	.	.	.	.
228	314	308	.	.	.	.	.
229	316	311	314	.	.	.	.
230	317	312	317	.	.	.	.
231	318	314	318	321	.	328	.
232	320	315	318	323	.	328	.
233	321	317	319	324	.	329	.
234	322	319	321	325	.	331	.
235	324	320	323	327	.	331	.
236	325	321	324	329	325	332	.
237	327	322	325	329	327	333	.
238	329	323	327	330	328	334	333
239	330	324	328	331	330	334	333
240	332	325	329	332	331	335	336
241	334	326	330	332	332	335	338
242	335	327	331	333	333	336	339
243	336	328	331	334	334	337	339
244	337	329	332	335	334	337	340
245	340	331	333	336	335	337	340
246	342	332	334	336	336	338	340
247	343	333	334	337	336	339	341
248	344	335	335	338	338	340	341
249	345	336	336	339	338	340	341
250	346	338	337	339	339	341	342
251	348	339	338	340	340	341	342
252	349	341	339	342	340	342	343

	Third Edition						
Second Edition	Grade 3P	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
253	351	343	340	343	341	343	343
254	352	344	341	344	342	344	344
255	355	346	342	345	343	345	344
256	356	348	344	346	344	346	345
257	359	348	345	347	345	347	346
258	359	350	346	348	346	347	346
259	362	351	348	349	347	348	347
260	.	353	349	350	348	350	348
261	.	354	350	351	349	350	349
262	.	356	352	352	350	351	349
263	.	357	353	353	351	352	350
264	.	359	354	354	352	353	351
265	.	360	355	355	353	354	352
266	.	362	356	356	354	355	353
267	.	363	357	357	355	356	354
268	.	364	358	358	356	357	355
269	.	365	359	359	356	357	355
270	.	366	360	360	357	358	356
271	.	367	362	361	358	359	357
272	.	368	363	362	360	360	358
273	.	.	364	363	361	361	359
274	.	.	365	364	362	361	360
275	.	.	367	365	363	362	361
276	.	.	368	366	364	363	362
277	.	.	369	367	365	364	363
278	.	.	370	368	366	365	364
279	.	.	372	369	366	365	365
280	.	.	374	369	367	366	366
281	.	.	374	370	368	367	367
282	.	.	374	372	370	367	367
283	.	.	.	373	371	368	368
284	.	.	.	374	371	369	369
285	.	.	.	374	372	370	370
286	.	.	.	374	374	370	371
287	.	.	.	374	375	371	372
288	.	.	.	375	376	372	373
289	.	.	.	376	378	373	373
290	.	.	.	377	380	374	374
291	.	.	.	.	383	374	375
292	.	.	.	.	383	375	376
293	.	.	.	.	383	376	377
294	.	.	.	.	.	376	377
295	.	.	.	.	.	377	379
296	.	.	.	.	.	377	379
297	.	.	.	.	.	377	380

	Third Edition						
Second Edition	**Grade 3P**	**Grade 3**	**Grade 4**	**Grade 5**	**Grade 6**	**Grade 7**	**Grade 8**
298	.	.	.	.	.	379	381
299	.	.	.	.	.	380	382
300	.	.	.	.	.	381	382
301	.	.	.	.	.	383	384
302	.	.	.	.	.	384	384
303	.	.	.	.	.	387	384
304	.	.	.	.	.	387	385
305	.	.	.	.	.	388	386

Setting Standards

Cut score ranges for the four achievement level categories developed for the Second Edition tests were applied to the Third Edition scores using the score translation table above (Table 6). These ranges appear in Table 7. In a few of the moderating data sets, no student scored as low as the Second Edition sample in Level I and no student scored as high as the Second Edition sample in Level IV. In addition, there were instances where the upper score of a level was equal to the lower score of the next higher level. Final cut score ranges were made in consultation with the North Carolina Department of Education.

Table 7. Cut scores for Second and Third Edition of the North Carolina EOG Tests of Mathematics scales for Grades 3P–8

Grade	Level	Second Edition	Third Edition
Pre 3	I	211–219	291–306
	II	220–229	307–316
	III	230–239	317–331
	IV	240–260	332–358
3	I	218–237	302–322
	II	238–245	323–331
	III	246–254	332–344
	IV	255–276	345–374
4	I	221–239	313–328
	II	240–246	329–334
	III	247–257	335–345
	IV	258–285	346–379
5	I	221–242	319–333
	II	243–249	334–339
	III	250–259	340–349
	IV	260–295	350–388
6	I	228–246	323–336
	II	247–253	337–341
	III	254–264	342–352
	IV	265–296	353–388
7	I	231–249	328–340
	II	250–257	341–346
	III	258–266	347–355
	IV	267–307	356–389
8	I	235–253	330–343
	II	254–260	344–348
	III	261–271	349–357
	IV	272–310	358–389

Quality Assurance Procedures

The authors have applied a variety of analyses and procedures to ensure that the results of the scaling and linking studies are correct. First, data sets for each grade and form were compiled from data sent by the North Carolina Department of Education. These data sets were scored by two independent researchers and the frequency distributions of raw total scores were compared. The frequency distributions were found to be identical. This process ensured that the scored data used in the analyses were accurate. Second, for the vertical scale, the mean difference and standard deviation ratios for the grades and forms were compared to the classical test theory *p*-values of the linking items and compared to Stocking-Lord (1983) equating values (slope and intercept) on IRT parameters provided separately by the North Carolina Department of Education. In both cases, the data provided evidence that the mean difference and standard deviation ratios were accurate in terms of direction and amount (see Tables 8 and 9). Third, the statistical method used to create the vertical scale was applied to the Second Edition data to ensure that it reproduced the scale correctly.

Table 8. Average mean difference in standard deviation units of the lower grade and standard deviation ratios, and average Stocking-Lord equating parameters from North Carolina-supplied IRT parameters, for the NC EOG Tests of Mathematics, Third Edition

Grade Pair	Average Mean Differences	Average Standard Deviation Ratios	Average Stocking Lord Equating Intercept	Average Stocking Lord Equating Slope
3P–3	0.982	0.864	1.015	0.943
3–4	0.531	0.933	0.520	0.917
4–5	0.463	0.977	0.453	0.965
5–6*	0.145	1.041	0.121	1.004
6–7	0.212	0.975	0.212	0.974
7–8	0.174	0.984	0.167	0.985

* One grade-pair form dropped as an outlier from analyses

Table 9. Average mean difference in standard deviation units of the lower grade and standard deviation ratios, and average differences in *p*-values (higher minus lower grade) of linking sets, for the North Carolina EOG Tests of Mathematics, Third Edition

Grade Pair	Average Mean Differences	Mean p-value Differences for Linking Items
3P–3	.982	.182
3–4	.531	.098
4–5	.463	.087
5–6*	.145	.032
6–7	.212	.040
7–8	.174	.030

* One grade-pair form dropped as an outlier from analyses

Psychometrics Underlying the Developmental Scale

The procedure for creating the developmental scale is based upon those described in Williams, Pommerich and Thissen (1998), and is divided into four steps. These steps are described below.

Step 1. The first step involved using BILOG-MG to calibrate the End-of-Grade math test item and population parameters for adjacent grades. This process was described in the section entitled "Third Edition Developmental Scale" of this report and resulted in an average mean difference and average standard deviation ratio (m_n and s_n) for each grade *n* (see Table 2).

Step 2. In the second step, a (0,1) growth scale anchored at Grade 3P was constructed to yield the following means (M_n) and standard deviations (S_n):

$M_n = M_{n-1} + m_n S_{n-1}$, mean for Grade *n* on (0,1) growth scale anchored at the lowest grade (with Grade 3P indexed as *n*=2),

$S_n = s_n S_{n-1}$, standard deviation for Grade *n* on (0,1) growth scale anchored at the lowest grade (with Grade 3P indexed as *n*=2),

where $M_2 \equiv 0$, and $S_2 \equiv 1$. This (0,1) growth scale was generated recursively upwards to Grade 8.

Step 3. In the third step, this scale was re-centered (re-anchored) at Grade 5 yielding,

$$M_n^* = \frac{(M_n - M_5)}{S_5}$$

$$S_n^* = \frac{S_n}{S_5}$$

as the means (M^*_n) and standard deviations (S^*_n).

Step 4. The fourth and final step in constructing the growth scale was the application of a linear transformation in order to produce a growth scale with the Grade 5 mean and standard deviations equal to 350 and 10, respectively, *viz.*

$$\mu_n = 350 + 10 M_n^*$$

$$\sigma_n = 10 S_n^* ,$$

where μ_n is the mean of the final growth scale in Grade *n* and σ_n is the standard deviation for the growth scale in Grade *n*.

References

Bazemore, M., Kramer, L., Yelton, A., & Brown, R. (2006). *North Carolina Mathematics Tests Technical Report.* Raleigh, NC: North Carolina Department of Public Instruction.

Kolen, M.J., & Brennan, R.L. (1995). *Test equating methods and practices.* New York: Springer.

Mislevy, R.J. (1992). *Linking educational assessments: Concepts, issues, methods, and prospects.* Princeton, NJ: Educational Testing Service.

Stocking, M.L., & Lord, F.M. (1983). Development of a common metric in Item Response Theory. *Applied Psychological Measurement, 7(2),* 201–210.

Thissen, D., Sathy, V., Edwards, M., & Flora, D. (2006). *North Carolina Mathematics Tests Technical Report, Appendix C.* Raleigh, NC: North Carolina Department of Public Instruction.

Williams, V.S.L., Pommerich, M., & Thissen, D. (1998). A comparison of developmental scales based upon Thurstone methods and item response theory. *Journal of Educational Measurement, 35,* 93–107.

Appendix D − Sample Items

Additional sample items can be accessed at the NCDPI website:
http://www.ncpublicschools.org/accountability/testing/eog/sampleitems

The following questions were developed for the North Carolina End-of-Grade (EOG) Tests and are aligned to the 2003 Mathematics *Standard Course of Study*. The North Carolina Department of Public Instruction (NCDPI) has chosen to release this group of questions so that the public may have access to questions similar to those that appear on the EOG tests administered during the school year 2005–2006 and beyond.

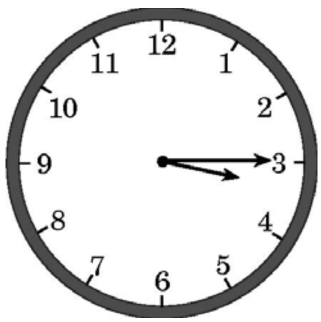
These sample questions reflect only a small part of the range of questions which might be asked to measure student achievement on the grade-level objectives of the SCS. While sample questions can be useful, teaching and learning should be driven by the full richness and content of the objectives.

Note: Drawings may not be to scale.

Grade 3 Goal 1
There are about ten bananas in one bunch. ***About*** how many bananas would be in 19 bunches?

A between 10 and 20
B between 80 and 100
C between 180 and 200
D between 1,800 and 2,000

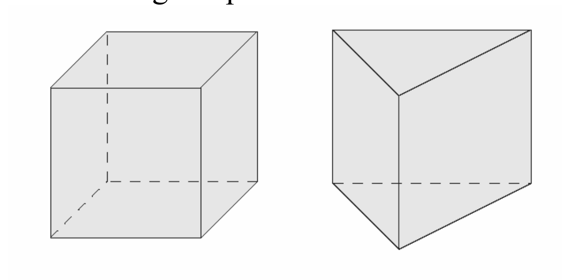
Grade 3 Goal 2



Marissa eats dinner at 6:30 p.m.
She gets home from school at 3:15 p.m. She practices soccer for 1 hour and 30 minutes, does chores for 45 minutes, and walks her dog for 20 minutes. Will she have at least 30 minutes before dinner to do her homework?

A No, after she gets home from school she only has 3 hours and 15 minutes before dinner.
B No, she spends too much time practicing soccer.
C Yes, she will have 40 minutes before dinner to do her homework.
D Yes, she will have one hour before dinner to do her homework.

Grade 3 Goal 3
How many more vertices does a cube have than a triangular prism?



A 8
B 6
C 4
D 2

Grade 3 Goal 4
Kyle has a pocket full of marbles. He has one black, two green, three red, and two yellow. He pulls out one marble. What is the likelihood that the marble will be black?

A 1 out of 1
B 1 out of 7
C 1 out of 8
D 7 out of 8

Grade 3 Goal 5
What is the rule for this number pattern?

1, 1, 2, 6, 24, 120, . . .

A add 0, then add 1, then add 2, and so on
B multiply by 1, then multiply by 2, then multiply by 3, and so on
C multiply by 1, then add 1
D multiply by two, then subtract 1

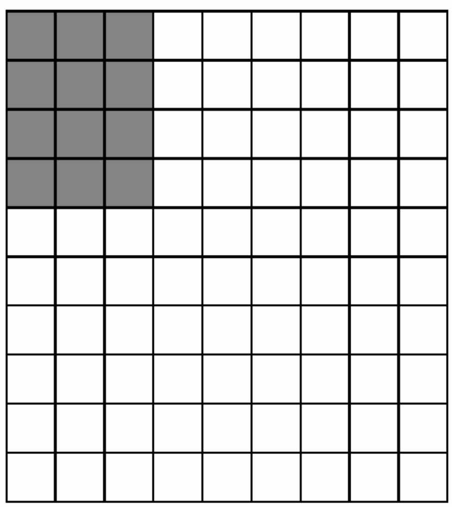
Grade 4 Goal 1

There are eighty-six thousand four hundred seconds in a day. How could this number be written?

A 80,064
B 80,640
C 86,400
D 86,404

Grade 4 Goal 2

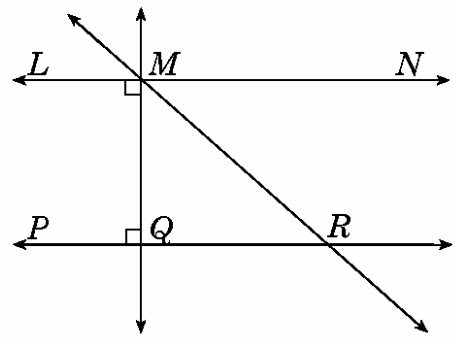
If the length and width of the shaded rectangle were doubled, what would be the area of the new rectangle?



A 60 square units
B 48 square units
C 36 square units
D 24 square units

Grade 4 Goal 3

Which line is parallel to $\overleftrightarrow{LN}$?



A $\overleftrightarrow{LM}$

B $\overleftrightarrow{MQ}$

C $\overleftrightarrow{MR}$

D $\overleftrightarrow{PR}$

Grade 4 Goal 4

The nurse recorded data on the number of people who donated blood in the last 10 community blood drives. The data is shown below.

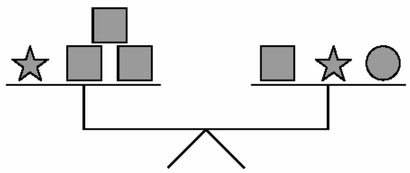
60, 64, 52, 62, 58, 57, 64, 64, 68, 52

What is the median number of donors the community has for blood drives?

A 10
B 16
C 61
D 64

Grade 4 Goal 5

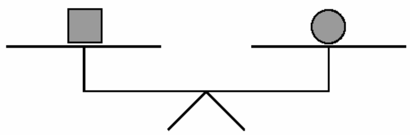
If



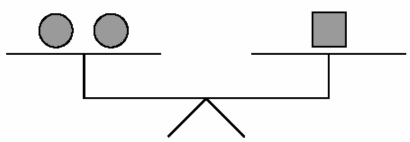
,

which of the following must be true?

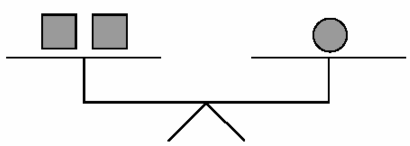
A



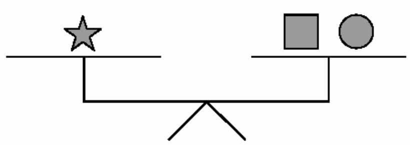
B



C



D



Grade 5 Goal 1

Sherry studied this group of fractions.

$$\frac{2}{3} \qquad \frac{2}{4} \qquad \frac{2}{5} \qquad \frac{2}{6}$$

What is true about the value of the fractions?

A Increasing the denominator increases the value of the fraction.

B If the denominator stays the same and the numerator increases, the fraction names a smaller amount.

C Increasing the denominator by adding 2 cuts the size of the fraction in half.

D If the numerator stays the same and the denominator increases, the fraction names a smaller amount.

Grade 5 Goal 2

Mrs. Aldridge has asked her students to predict how many liters of water it will take to fill a 10-quart bucket. Which prediction is close to 10 quarts?

A 80 liters

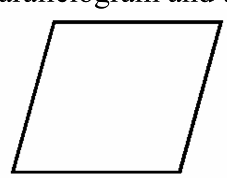
B 40 liters

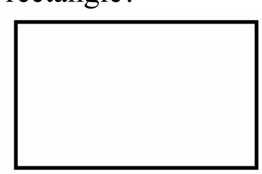
C 20 liters

D 10 liters

Grade 5 Goal 3

Which statement correctly compares this parallelogram and this rectangle?





Parallelogram Rectangle

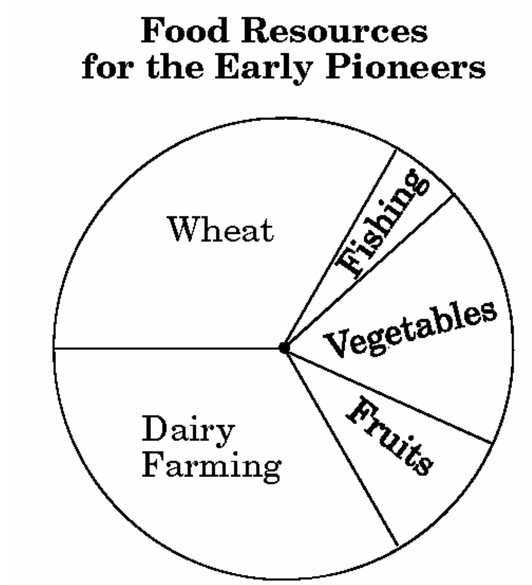
A Both figures are polygons with pairs of parallel opposite sides.

B Both figures are polygons with four right angles.

C Both figures are polygons whose interior angles total 180 degrees.

D Both figures are polygons with at least two acute angles.

Grade 5 Goal 4
According to this graph, which of the following is true about the early pioneers' food resources?



A Fishing and fruits combined would be the largest food resources.
B Wheat and dairy farming make up about the same amount of the food resources.
C Fruits provide the least amount of food resources.
D Vegetables and dairy farming make up the same amount of the food resources.

Grade 5 Goal 5
There are 36 fifth-graders in art class. The art teacher wants to arrange their pictures on the wall so that one is in the first row, two are in the second row, three are in the third row, etc. How many rows of pictures will there be?
A 8 rows
B 9 rows
C 10 rows
D 11 rows

Grade 6 Goal 1
The radius of an atom is one nanometer, which is approximately 3.937×10^{-8} inch. What is this length expressed in standard notation?
A 0.000000003937 in.
B 0.00000003937 in.
C 0.0000003937 in.
D 0.000003937 in.

Grade 6 Goal 2
The diameter of a jar lid is 5.4 cm. What is the ***approximate*** area of the top of the lid?
A 9 cm^2
B 17 cm^2
C 23 cm^2
D 92 cm^2

Grade 6 Goal 3
A line and a triangle are in the same plane. The line intersects the triangle at exactly one point, *P*. Which statement is true?
A *P* is a vertex of the triangle.
B *P* is a midpoint of a side of the triangle.
C *P* is in the interior of the triangle.
D *P* is in the exterior of the triangle.

Grade 6 Goal 4
Matt has a bag containing 12 green marbles and 8 blue marbles. Without looking, he pulls out one marble and places it on a table. He then picks a second marble from the bag. What is the probability he will have 2 blue marbles?
A $\frac{8}{20} \bullet \frac{7}{19}$
B $\frac{8}{20} \bullet \frac{7}{20}$
C $\frac{1}{8} \bullet \frac{1}{7}$
D $\frac{1}{8} \bullet \frac{1}{8}$

Grade 6 Goal 5

Evaluate: $2 + (8 - 4) + 3^2 \times \frac{6}{2}$

A 36

B 35

C 33

D 24

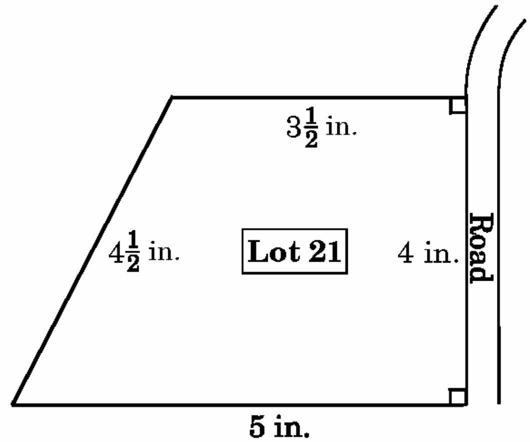
Grade 7 Goal 1

A 6-ounce box of strawberry gelatin costs \$0.90 and a 3-ounce box costs \$0.42. What is the difference in cost per ounce between the larger and the smaller boxes?

A \$0.48
B \$0.04
C \$0.03
D \$0.01

Grade 7 Goal 2

Lot 21 is a trapezoid with the two bases perpendicular to the road. The scale drawing below uses the scale $\frac{1}{2}$ inch = 40 feet.



What is the ***approximate*** area of Lot 21? $\left[A = \frac{1}{2}h(b_1 + b_2)\right]$
(1 acre = 43,560 square feet)

A $\frac{1}{2}$ acre

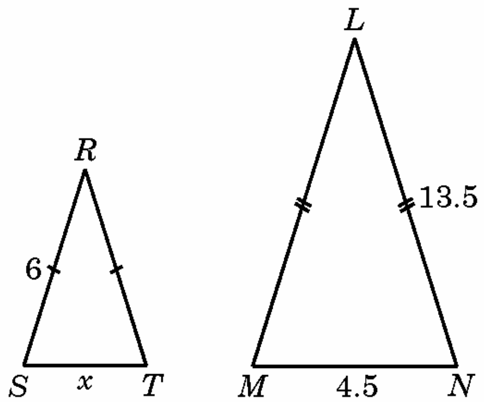
B 2 acres

C $2\frac{1}{2}$ acres

D 3 acres

Grade 7 Goal 3

ΔRST is an isosceles triangle with $\overline{RS} \cong \overline{RT}$.



If ΔRST is similar to ΔLMN, what is the value of x?

A 2

B 3

C 4

D 5

Grade 7 Goal 4

What is the ***approximate*** mean for this set of temperatures?

92°F, 89°F, 90°F, 78°F, 83°F, 90°F, 88°F

A 86.7°F
B 87.1°F
C 87.4°F
D 88.7°F

Grade 7 Goal 5

What is the ***approximate*** volume of a cone with a radius of 6 inches and a height of 9 inches? $V = \frac{1}{3}\pi r^2 h$

A 113.04 in.^{3}
B 339.29 in.^{3}
C 1,017.88 in.^{3}
D 1,356.48 in.^{3}

Grade 8 Goal 1

The drama club is selling tickets to a play for $10 each. The cost to rent the theater and costumes is $500. In addition, the printers are charging $1 per ticket to print the tickets. How many tickets must the drama club sell to make a profit?

A 54
B 55
C 56
D 57

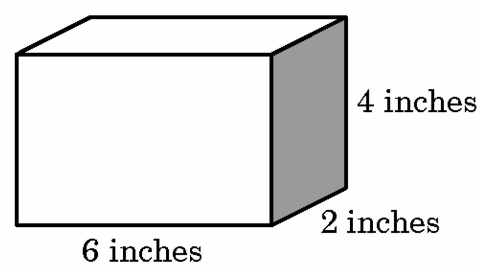
Grade 8 Goal 2

At noon, the shadow of a flagpole is 19 feet long. At the same time, the shadow of a 12-foot-high wall is 4 feet long. What is the height of the flagpole?

A 48 feet
B 57 feet
C 62 feet
D 75 feet

Grade 8 Goal 3

What is the maximum number of cubes with a side length of 2 inches that can fit in this box?



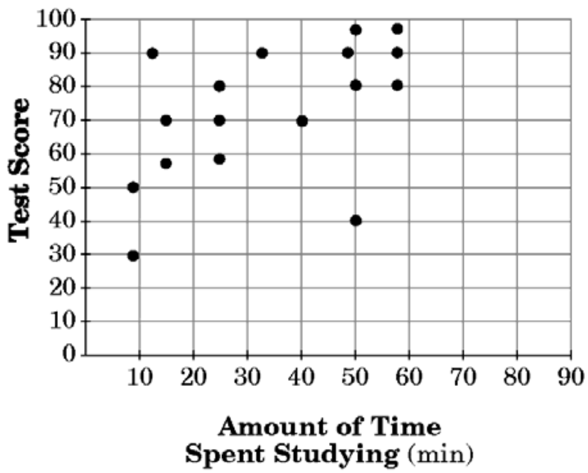
A 48

B 24

C 12

D 6

Grade 8 Goal 4
Which relationship is suggested by the scatterplot below?



A The amount of time spent studying does not affect a test score.
B the longer amount of time spent studying, the higher the test score
C the longer amount of time spent studying, the lower the test score
D the shorter amount of time spent studying, the higher the test score

Grade 8 Goal 5
The price of a large pizza is given by the formula $P(t) = 1.5t + 7.50$, where is the price of the pizza and is the number of toppings. What does the slope represent?
A number of toppings
B cost per slice
C cost of each topping
D cost of the pizza with no toppings

Algebra I Goal 1
To find the image length, L, of a 4-foot-tall object in a spherical mirror with a focal length of 2 feet,

$$L = 4\left(\frac{2}{o-2}\right)^2$$

can be used, where o is the distance, in feet, of the object from the mirror. What is the image length of the object when it is 1.5 feet away from the mirror?
A 256 feet
B 128 feet
C 64 feet
D 32 feet

Algebra I Goal 2
The equation of the line containing one side of a parallelogram is $3x + 2y = 8$. The opposite side contains the point $(0, {}^{-}7)$. Which is the equation of the line that contains the opposite side?

A $y = \frac{2}{3}x - 7$

B $y = {}^{-}\frac{3}{2}x + 7$

C $y = \frac{2}{3}x + 7$

D $y = {}^{-}\frac{3}{2}x - 7$

Algebra I Goal 3

This matrix shows the cost of cell phone service offered by several different companies.

	Monthly Cost for 200 Minutes	Cost of Each Minute over 200 Minutes
Company 1	$39.00	$0.05
Company 2	$27.00	$0.08
Company 3	$42.00	$0.04
Company 4	$30.00	$0.06

What is the cost of 320 minutes with Company 4?

A $37.20
B $45.00
C $49.20
D $75.00

Algebra I Goal 4

When Robert was born, his grandfather invested $1,000 for Robert's college education. At an interest rate of 4.5%, compounded annually, ***approximately*** how much would Robert have at age 18?
(use the formula $A = P(1 + r)^t$ where P is the principal, r is the interest rate, and t is the time in years)

A $1,810
B $2,200
C $3,680
D $18,810

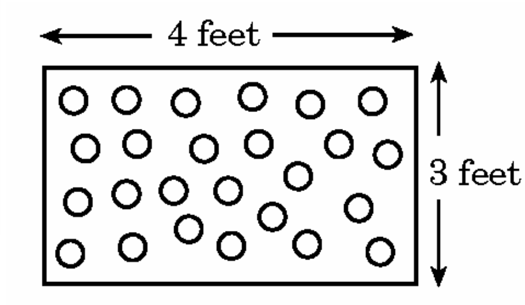
Geometry Goal 1

An inflated round balloon with radius $r = 50$ centimeters holds approximately 523,600 cubic centimeters of air. When the balloon is contracted such that the radius is ⅔ the original size, what is the ***approximate*** volume of the partially deflated balloon?

A 1.94×10^4 cm^3
B 1.55×10^5 cm^3
C 1.75×10^5 cm^3
D 3.49×10^5 cm^3

Geometry Goal 1

To win a carnival game, Keisha must throw a dart and hit one of 25 circles in a dart board that is 4 feet by 3 feet. The diameter of each circle is 4 inches.

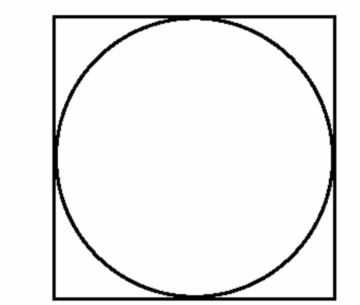


Approximately what is the probability that a randomly thrown dart that hits the board would also hit a circle?

A 18%
B 26%
C 63%
D 73%

Geometry Goal 2

A gardener wants to enclose a circular garden with a square fence, as shown below.



If the circumference of the circular garden is about 48 feet, which of the following is the ***approximate*** length of fencing needed?

A 31 ft
B 61 ft
C 122 ft
D 244 ft

Geometry Goal 3

$\Delta P'Q'R'$ is the image produced after reflecting ΔPQR across the y-axis. If vertex P has coordinates (s, t), what are the coordinates of P'?

A (t, s)

B $(s, {}^{-}t)$

C $({}^{-}s, {}^{-}t)$

D $({}^{-}s, t)$

Algebra II Goal 1

The amount of simple interest earned on a savings account varies jointly with time, *t* (in years), and the principal, *p* (in dollars). After 5 years, interest on $800 in this savings account is $260.00. What is the annual interest rate (constant of variation)?

A 7.4%
B 6.5%
C 5.4%
D 2.7%

Algebra II Goal 2

A car insurance company has a special plan for safe drivers. For each year that a driver has no tickets or violations, the premium is reduced by 10%, and a credit of $15.00 is awarded. Which equation shows the amount a driver with no tickets or violations owes in the $(n + 1)$th year as a function of the amount owed in the nth year?

A $f(n + 1) = f(n) - 0.10\, f(n) - 15$
B $f(n + 1) = f(n) + 0.10\, f(n) + 15$
C $f(n + 1) = f(n) + 0.10\, f(n) - 15$
D $f(n + 1) = f(n) - 0.10\, f(n) + 15$

Algebra II Goal 2

A ball is tossed from the top of a building. This table shows the height, *h* (in feet), of the ball above the ground *t* seconds after being tossed.

t	1	2	3	4	5	6
h	299	311	291	239	155	39

According to a quadratic best-fit model of the data, how long after the ball was tossed was it 80 feet above the ground?

A about 5.1 seconds
B about 5.4 seconds
C about 5.7 seconds
D about 5.9 seconds

Appendix E – Frequency Distribution Tables for Math Scale Scores

State North Carolina End-of-Grade Testing Program

Grade 3 Mathematics Pretest 2006–2007

Summary Statistics on Developmental Scale Scores

Number of Students with Valid Scores	108501	High Score	358
		Low Score	293
Mean	329.7	State Percentiles	Scale Score
Standard Deviation	11.4	90	344.31
		75	337.72
		50 (Median)	330.04
Variance	128.9	25	321.47
		10	314.22

Frequency Distribution

Developmental Scale Score	Frequency	Cumulative Frequency	Percent	Cumulative Percent	2006–07 State Percentile
358	578	108501	0.53	100.00	99
354	1311	107923	1.21	99.47	99
351	1802	106612	1.66	98.26	97
350	355	104810	0.33	96.60	96
348	2803	104455	2.58	96.27	95
346	3422	101652	3.15	93.69	92
344	3080	98230	2.84	90.53	89
343	611	95150	0.56	87.70	87
342	4078	94539	3.76	87.13	85
341	780	90461	0.72	83.37	83
340	3461	89681	3.19	82.65	81
339	3646	86220	3.36	79.46	78
338	1535	82574	1.41	76.10	75
337	3682	81039	3.39	74.69	73
336	3883	77357	3.58	71.30	70
335	4448	73474	4.10	67.72	66
334	1558	69026	1.44	63.62	63
333	3673	67468	3.39	62.18	60
332	3780	63795	3.48	58.80	57
331	4424	60015	4.08	55.31	53
330	2906	55591	2.68	51.24	50
329	2210	52685	2.04	48.56	48
328	3624	50475	3.34	46.52	45
327	4338	46851	4.00	43.18	41
326	3541	42513	3.26	39.18	38
325	2884	38972	2.66	35.92	35
324	2092	36088	1.93	33.26	32
323	3322	33996	3.06	31.33	30
322	3461	30674	3.19	28.27	27
321	2811	27213	2.59	25.08	24
320	1884	24402	1.74	22.49	22
319	3876	22518	3.57	20.75	19
318	1903	18642	1.75	17.18	16
317	1669	16739	1.54	15.43	15
316	2868	15070	2.64	13.89	13
315	542	12202	0.50	11.25	11
314	2914	11660	2.69	10.75	9
313	963	8746	0.89	8.06	8
312	1595	7783	1.47	7.17	6
311	1603	6188	1.48	5.70	5
310	347	4585	0.32	4.23	4
309	1314	4238	1.21	3.91	3
308	225	2924	0.21	2.69	3
307	1134	2699	1.05	2.49	2
306	241	1565	0.22	1.44	1
305	485	1324	0.45	1.22	1

304	235	839	0.22	0.77	1
303	182	604	0.17	0.56	1
302	185	422	0.17	0.39	1
301	32	237	0.03	0.22	1
300	82	205	0.08	0.19	1
299	32	123	0.03	0.11	1
298	13	91	0.01	0.08	1
297	13	78	0.01	0.07	1
296	20	65	0.02	0.06	1
295	8	45	0.01	0.04	1
Less than 295	37	37	0.03	0.03	1

PUBLIC SCHOOLS OF NORTH CAROLINA END-OF-GRADE TESTS 2005–2006
GRADE 3 MATH DEVELOPMENTAL SCALE SCORE FREQUENCY REPORT
Regular test administration

State Results

SUMMARY STATISTICS

Number of Students with Valid Scores	104281	High Score	370
		Low Score	311
		Local Percentiles	Dev Scale Score
Mean	343.20	90	356.00
		75	350.00
Standard Deviation	9.70	50 (Median)	343.00
		25	337.00
Mode	343	10	330.00

FREQUENCY DISTRIBUTION

DEV SCALE SCORE	FREQUENCY	CUMULATIVE FREQUENCY	PERCENT	CUMULATIVE PERCENT	ACHIEVEMENT LEVEL
370	57	104281	0.05	100.00	IV
369	190	104224	0.18	99.95	IV
366	661	104034	0.63	99.76	IV
364	641	103373	0.61	99.13	IV
363	348	102732	0.33	98.51	IV
362	844	102384	0.81	98.18	IV
361	470	101540	0.45	97.37	IV
360	1658	101070	1.59	96.92	IV
359	290	99412	0.28	95.33	IV
358	1649	99122	1.58	95.05	IV
357	2171	97473	2.08	93.47	IV
356	1984	95302	1.90	91.39	IV
355	2085	93318	2.00	89.49	IV
354	1830	91233	1.75	87.49	IV
353	2927	89403	2.81	85.73	IV
352	3044	86476	2.92	82.93	IV
351	3121	83432	2.99	80.01	IV
350	4456	80311	4.27	77.01	IV
349	3342	75855	3.20	72.74	IV
348	4485	72513	4.30	69.54	IV
347	3508	68028	3.36	65.24	IV
346	3593	64520	3.45	61.87	IV
345	3574	60927	3.43	58.43	IV
344	4752	57353	4.56	55.00	IV
343	5485	52601	5.26	50.44	IV
342	4188	47116	4.02	45.18	IV
341	3477	42928	3.33	41.17	IV
340	3554	39451	3.41	37.83	IV
339	3301	35897	3.17	34.42	IV
338	3305	32596	3.17	31.26	IV
337	3242	29291	3.11	28.09	IV
336	3206	26049	3.07	24.98	IV
335	2904	22843	2.78	21.91	IV
334	2333	19939	2.24	19.12	IV
333	2610	17606	2.50	16.88	IV
332	2060	14996	1.98	14.38	IV
331	1533	12936	1.47	12.40	IV

330	1887	11403	1.81	10.93	IV
329	1714	9516	1.64	9.13	IV
328	1536	7802	1.47	7.48	IV
327	848	6266	0.81	6.01	IV
326	1273	5418	1.22	5.20	IV
325	851	4145	0.82	3.97	IV
324	1042	3294	1.00	3.16	IV
323	446	2252	0.43	2.16	IV
322	600	1806	0.58	1.73	IV
321	389	1206	0.37	1.16	IV
320	357	817	0.34	0.78	IV
319	169	460	0.16	0.44	IV
318	148	291	0.14	0.28	IV
317	91	143	0.09	0.14	IV
316	29	52	0.03	0.05	IV
315	10	23	0.01	0.02	IV
314	2	13	0.00	0.01	IV
313	3	11	0.00	0.01	IV
312	6	8	0.01	0.01	IV
311	2	2	0.00	0.00	IV

PUBLIC SCHOOLS OF NORTH CAROLINA END-OF-GRADE TESTS 2005–2006
GRADE 4 MATH DEVELOPMENTAL SCALE SCORE FREQUENCY REPORT
Regular test administration

STATE LEVEL

SUMMARY STATISTICS

Number of Students with Valid Scores	102366	High Score	374
		Low Score	319
		Local Percentiles	Dev Scale Score
Mean	348.90	90	361.00
		75	356.00
Standard Deviation	9.46	50 (Median)	349.00
		25	342.00
		10	336.00

FREQUENCY DISTRIBUTION

DEV SCALE SCORE	FREQUENCY	CUMULATIVE FREQUENCY	PERCENT	CUMULATIVE PERCENT	ACHIEVEMENT LEVEL
374	47	102366	0.05	100.00	IV
373	307	102319	0.30	99.95	IV
372	58	102012	0.06	99.65	IV
371	140	101954	0.14	99.60	IV
370	838	101814	0.82	99.46	IV
368	446	100976	0.44	98.64	IV
367	1064	100530	1.04	98.21	IV
366	851	99466	0.83	97.17	IV
365	1256	98615	1.23	96.34	IV
364	1695	97359	1.66	95.11	IV
363	1138	95664	1.11	93.45	IV
362	1964	94526	1.92	92.34	IV
361	1998	92562	1.95	90.42	IV
360	2536	90564	2.48	88.47	IV
359	2242	88028	2.19	85.99	IV
358	3581	85786	3.50	83.80	IV
357	2761	82205	2.70	80.30	IV
356	3328	79444	3.25	77.61	IV
355	3915	76116	3.82	74.36	IV
354	3291	72201	3.21	70.53	IV
353	3462	68910	3.38	67.32	IV
352	5003	65448	4.89	63.94	IV
351	3519	60445	3.44	59.05	IV
350	3117	56926	3.04	55.61	IV
349	4587	53809	4.48	52.57	IV
348	4766	49222	4.66	48.08	IV
347	3197	44456	3.12	43.43	IV
346	3176	41259	3.10	40.31	IV
345	3171	38083	3.10	37.20	IV
344	4680	34912	4.57	34.11	IV
343	3764	30232	3.68	29.53	IV
342	2666	26468	2.60	25.86	IV
341	3070	23802	3.00	23.25	IV
340	3072	20732	3.00	20.25	IV

339	2372	17660	2.32	17.25	IV
338	2223	15288	2.17	14.93	IV
337	2135	13065	2.09	12.76	IV
336	2298	10930	2.24	10.68	IV
335	1675	8632	1.64	8.43	IV
334	1647	6957	1.61	6.80	IV
333	793	5310	0.77	5.19	IV
332	1383	4517	1.35	4.41	IV
331	885	3134	0.86	3.06	IV
330	788	2249	0.77	2.20	IV
328	330	1051	0.32	1.03	IV
327	352	721	0.34	0.70	IV
326	208	369	0.20	0.36	IV
325	91	161	0.09	0.16	IV
324	40	70	0.04	0.07	IV
323	15	30	0.01	0.03	IV
322	6	15	0.01	0.01	IV
321	5	9	0.00	0.01	IV
319	4	4	0.00	0.00	IV

PUBLIC SCHOOLS OF NORTH CAROLINA END-OF-GRADE TESTS 2005–2006
GRADE 5 MATH DEVELOPMENTAL SCALE SCORE FREQUENCY REPORT
Regular test administration

State Results

SUMMARY STATISTICS

Number of Students with Valid Scores	103127	High Score	378
		Low Score	326
		Local Percentiles	Dev Scale Score
Mean	353.74	90	366.00
		75	360.00
Standard Deviation	9.25	50 (Median)	354.00
		25	347.00
Mode	352	10	341.00

FREQUENCY DISTRIBUTION

DEV SCALE SCORE	FREQUENCY	CUMULATIVE FREQUENCY	PERCENT	CUMULATIVE PERCENT	ACHIEVEMENT LEVEL
378	32	103127	0.03	100.00	IV
377	405	103095	0.39	99.97	IV
375	609	102690	0.59	99.58	IV
374	373	102081	0.36	98.99	IV
373	173	101708	0.17	98.62	IV
372	1192	101535	1.16	98.46	IV
371	250	100343	0.24	97.30	IV
370	1514	100093	1.47	97.06	IV
369	1312	98579	1.27	95.59	IV
368	1137	97267	1.10	94.32	IV
367	1855	96130	1.80	93.22	IV
366	1945	94275	1.89	91.42	IV
365	2558	92330	2.48	89.53	IV
364	2638	89772	2.56	87.05	IV
363	2736	87134	2.65	84.49	IV
362	3209	84398	3.11	81.84	IV
361	3043	81189	2.95	78.73	IV
360	3984	78146	3.86	75.78	IV
359	2963	74162	2.87	71.91	IV
358	5114	71199	4.96	69.04	IV
357	3133	66085	3.04	64.08	IV
356	4710	62952	4.57	61.04	IV
355	4821	58242	4.67	56.48	IV
354	3189	53421	3.09	51.80	IV
353	3752	50232	3.64	48.71	IV
352	6016	46480	5.83	45.07	IV
351	3249	40464	3.15	39.24	IV
350	3330	37215	3.23	36.09	IV
349	3609	33885	3.50	32.86	IV
348	3051	30276	2.96	29.36	IV
347	3571	27225	3.46	26.40	IV
346	2950	23654	2.86	22.94	IV
345	2356	20704	2.28	20.08	IV
344	2750	18348	2.67	17.79	IV

343	2212	15598	2.14	15.13	IV
342	2558	13386	2.48	12.98	IV
341	2290	10828	2.22	10.50	IV
340	1375	8538	1.33	8.28	IV
339	1867	7163	1.81	6.95	IV
338	774	5296	0.75	5.14	IV
337	1395	4522	1.35	4.38	IV
336	977	3127	0.95	3.03	IV
335	849	2150	0.82	2.08	IV
334	509	1301	0.49	1.26	IV
333	372	792	0.36	0.77	IV
332	221	420	0.21	0.41	IV
331	145	199	0.14	0.19	IV
330	38	54	0.04	0.05	IV
329	8	16	0.01	0.02	IV
328	2	8	0.00	0.01	IV
327	4	6	0.00	0.01	IV
326	2	2	0.00	0.00	IV

PUBLIC SCHOOLS OF NORTH CAROLINA END-OF-GRADE TESTS 2005–2006
GRADE 6 MATH DEVELOPMENTAL SCALE SCORE FREQUENCY REPORT
Regular test administration

State Results

SUMMARY STATISTICS

Number of Students with Valid Scores	106141	High Score	381
		Low Score	328
		Local Percentiles	Dev Scale Score
Mean	354.91	90	368.00
		75	362.00
Standard Deviation	9.70	50 (Median)	355.00
		25	348.00
Mode	355	10	342.00

FREQUENCY DISTRIBUTION

DEV SCALE SCORE	FREQUENCY	CUMULATIVE FREQUENCY	PERCENT	CUMULATIVE PERCENT	ACHIEVEMENT LEVEL
381	79	106141	0.07	100.00	IV
380	317	106062	0.30	99.93	IV
378	223	105745	0.21	99.63	IV
377	434	105522	0.41	99.42	IV
376	125	105088	0.12	99.01	IV
375	906	104963	0.85	98.89	IV
374	193	104057	0.18	98.04	IV
373	1170	103864	1.10	97.85	IV
372	794	102694	0.75	96.75	IV
371	1427	101900	1.34	96.00	IV
370	1557	100473	1.47	94.66	IV
369	1959	98916	1.85	93.19	IV
368	2136	96957	2.01	91.35	IV
367	2211	94821	2.08	89.33	IV
366	2276	92610	2.14	87.25	IV
365	2382	90334	2.24	85.11	IV
364	2922	87952	2.75	82.86	IV
363	3389	85030	3.19	80.11	IV
362	4002	81641	3.77	76.92	IV
361	2783	77639	2.62	73.15	IV
360	3849	74856	3.63	70.53	IV
359	4303	71007	4.05	66.90	IV
358	3732	66704	3.52	62.84	IV
357	3108	62972	2.93	59.33	IV
356	4378	59864	4.12	56.40	IV
355	5566	55486	5.24	52.28	IV
354	3390	49920	3.19	47.03	IV
353	3355	46530	3.16	43.84	IV
352	3358	43175	3.16	40.68	IV
351	3381	39817	3.19	37.51	IV
350	3534	36436	3.33	34.33	IV
349	3489	32902	3.29	31.00	IV
348	3559	29413	3.35	27.71	IV
347	3431	25854	3.23	24.36	IV

346	3379	22423	3.18	21.13	IV
345	2209	19044	2.08	17.94	IV
344	2092	16835	1.97	15.86	IV
343	2067	14743	1.95	13.89	IV
342	2833	12676	2.67	11.94	IV
341	2089	9843	1.97	9.27	IV
340	1524	7754	1.44	7.31	IV
339	1331	6230	1.25	5.87	IV
338	1427	4899	1.34	4.62	IV
337	1338	3472	1.26	3.27	IV
336	907	2134	0.85	2.01	IV
335	483	1227	0.46	1.16	IV
334	364	744	0.34	0.70	IV
333	206	380	0.19	0.36	IV
332	107	174	0.10	0.16	IV
331	41	67	0.04	0.06	IV
330	20	26	0.02	0.02	IV
329	4	6	0.00	0.01	IV
328	2	2	0.00	0.00	IV

PUBLIC SCHOOLS OF NORTH CAROLINA END-OF-GRADE TESTS 2005–2006
GRADE 7 MATH DEVELOPMENTAL SCALE SCORE FREQUENCY REPORT
Regular test administration

State Results

SUMMARY STATISTICS

Number of Students with Valid Scores	105930	High Score	383
		Low Score	332
Mean	357.76	Local Percentiles	Dev Scale Score
		90	371.00
		75	365.00
Standard Deviation	9.65	50 (Median)	358.00
		25	351.00
Mode	356	10	345.00

FREQUENCY DISTRIBUTION

DEV SCALE SCORE	FREQUENCY	CUMULATIVE FREQUENCY	PERCENT	CUMULATIVE PERCENT	ACHIEVEMENT LEVEL
383	80	105930	0.08	100.00	IV
382	581	105850	0.55	99.92	IV
380	191	105269	0.18	99.38	IV
379	790	105078	0.75	99.20	IV
378	502	104288	0.47	98.45	IV
377	791	103786	0.75	97.98	IV
376	774	102995	0.73	97.23	IV
375	1223	102221	1.15	96.50	IV
374	977	100998	0.92	95.34	IV
373	1697	100021	1.60	94.42	IV
372	1380	98324	1.30	92.82	IV
371	2219	96944	2.09	91.52	IV
370	1878	94725	1.77	89.42	IV
369	2288	92847	2.16	87.65	IV
368	2761	90559	2.61	85.49	IV
367	2469	87798	2.33	82.88	IV
366	3310	85329	3.12	80.55	IV
365	2893	82019	2.73	77.43	IV
364	3290	79126	3.11	74.70	IV
363	3631	75836	3.43	71.59	IV
362	3474	72205	3.28	68.16	IV
361	4210	68731	3.97	64.88	IV
360	3186	64521	3.01	60.91	IV
359	4855	61335	4.58	57.90	IV
358	3893	56480	3.68	53.32	IV
357	3029	52587	2.86	49.64	IV
356	6412	49558	6.05	46.78	IV
355	3216	43146	3.04	40.73	IV
354	3199	39930	3.02	37.69	IV
353	3341	36731	3.15	34.67	IV
352	3264	33390	3.08	31.52	IV
351	3868	30126	3.65	28.44	IV
350	3246	26258	3.06	24.79	IV
349	3375	23012	3.19	21.72	IV

348	2759	19637	2.60	18.54	IV
347	3209	16878	3.03	15.93	IV
346	1890	13669	1.78	12.90	IV
345	2277	11779	2.15	11.12	IV
344	1779	9502	1.68	8.97	IV
343	1799	7723	1.70	7.29	IV
342	1863	5924	1.76	5.59	IV
341	998	4061	0.94	3.83	IV
340	1169	3063	1.10	2.89	IV
339	682	1894	0.64	1.79	IV
338	593	1212	0.56	1.14	IV
337	358	619	0.34	0.58	IV
336	169	261	0.16	0.25	IV
335	69	92	0.07	0.09	IV
334	13	23	0.01	0.02	IV
332	10	10	0.01	0.01	IV

PUBLIC SCHOOLS OF NORTH CAROLINA END-OF-GRADE TESTS 2005–2006
GRADE 8 MATH DEVELOPMENTAL SCALE SCORE FREQUENCY REPORT
Regular test administration

State Results

SUMMARY STATISTICS

Number of Students with Valid Scores	107171	High Score	384
		Low Score	332
Mean	359.15	Local Percentiles	Dev Scale Score
		90	371.00
		75	366.00
Standard Deviation	9.21	50 (Median)	359.00
		25	353.00
Mode	360	10	347.00

FREQUENCY DISTRIBUTION

DEV SCALE SCORE	FREQUENCY	CUMULATIVE FREQUENCY	PERCENT	CUMULATIVE PERCENT	ACHIEVEMENT LEVEL
384	137	107171	0.13	100.00	IV
383	219	107034	0.20	99.87	IV
382	277	106815	0.26	99.67	IV
381	187	106538	0.17	99.41	IV
380	197	106351	0.18	99.23	IV.
379	472	106154	0.44	99.05	IV
378	1203	105682	1.12	98.61	IV
377	331	104479	0.31	97.49	IV
376	918	104148	0.86	97.18	IV
375	1461	103230	1.36	96.32	IV
374	1534	101769	1.43	94.96	IV
373	1628	100235	1.52	93.53	IV
372	2001	98607	1.87	92.01	IV
371	1749	96606	1.63	90.14	IV
370	2244	94857	2.09	88.51	IV
369	3283	92613	3.06	86.42	IV
368	2870	89330	2.68	83.35	IV
367	3028	86460	2.83	80.67	IV
366	3718	83432	3.47	77.85	IV
365	3502	79714	3.27	74.38	IV
364	4252	76212	3.97	71.11	IV
363	3779	71960	3.53	67.15	IV
362	4549	68181	4.24	63.62	IV
361	3482	63632	3.25	59.37	IV
360	5918	60150	5.52	56.13	IV
359	3091	54232	2.88	50.60	IV
358	5580	51141	5.21	47.72	IV
357	3915	45561	3.65	42.51	IV
356	3332	41646	3.11	38.86	IV
355	4157	38314	3.88	35.75	IV
354	3273	34157	3.05	31.87	IV
353	4142	30884	3.86	28.82	IV
352	3310	26742	3.09	24.95	IV
351	3175	23432	2.96	21.86	IV
350	2282	20257	2.13	18.90	IV

349	2974	17975	2.78	16.77	IV
348	2706	15001	2.52	14.00	IV
347	2408	12295	2.25	11.47	IV
346	1646	9887	1.54	9.23	IV
345	1965	8241	1.83	7.69	IV
344	1677	6276	1.56	5.86	IV
343	1439	4599	1.34	4.29	IV
342	1081	3160	1.01	2.95	IV
341	741	2079	0.69	1.94	IV
340	649	1338	0.61	1.25	IV
339	324	689	0.30	0.64	IV
338	202	365	0.19	0.34	IV
337	122	163	0.11	0.15	IV
336	22	41	0.02	0.04	IV
335	5	19	0.00	0.02	IV
334	4	14	0.00	0.01	IV
333	7	10	0.01	0.01	IV
332	3	3	0.00	0.00	IV

State North Caroline End-Of-Course Testing Program
Algebra I—–2007

Summary Statistics on Scale Scores

Number of Students With Valid Scores	116209	High Score	181
		Low Score	118
Mean	150.9	State Percentiles	Scale Score
Standard Deviation	10.2	90	164.3
		75	157.8
		50 (Median)	150.7
Variance	103.5	25	143.5
		10	137.5

Frequency Distribution

Scale Score	Frequency	Cumulative Frequency	Percent	Cumulative Percent	2006-07 State Percentiles
181	24	116209	0.02	100.00	99
180	44	116185	0.04	99.98	99
179	143	116141	0.12	99.94	99
178	153	115998	0.13	99.82	99
177	346	115845	0.30	99.69	99
176	283	115499	0.24	99.39	99
175	458	115216	0.39	99.15	99
174	120	114758	0.10	98.75	99
173	947	114638	0.81	98.65	98
172	486	113691	0.42	97.83	98
171	716	113205	0.62	97.42	97
170	1121	112489	0.96	96.80	96
169	811	111368	0.70	95.83	95
168	1227	110557	1.06	95.14	95
167	1221	109330	1.05	94.08	94
166	1712	108109	1.47	93.03	92
165	1496	106397	1.29	91.56	91
164	2062	104901	1.77	90.27	89
163	2211	102839	1.90	88.49	88
162	2228	100628	1.92	86.59	86
161	3403	98400	2.93	84.68	83
160	2631	94997	2.26	81.75	81
159	3222	92366	2.77	79.48	78
158	2831	89144	2.44	76.71	75
157	2943	86313	2.53	74.27	73
156	4971	83370	4.28	71.74	70
155	4478	78399	3.85	67.46	66
154	3251	73921	2.80	63.61	62
153	3299	70670	2.84	60.81	59
152	4668	67371	4.02	57.97	56
151	5600	62703	4.82	53.96	52
150	3519	57103	3.03	49.14	48
149	4053	53584	3.49	46.11	44
148	4095	49531	3.52	42.62	41
147	5416	45436	4.66	39.10	37
146	3152	40020	2.71	34.44	33
145	3679	36868	3.17	31.73	30
144	4169	33189	3.59	28.56	27
143	4007	29020	3.45	24.97	23
142	2758	25013	2.37	21.52	20
141	3641	22255	3.13	19.15	18
140	2411	18614	2.07	16.02	15
139	2320	16203	2.00	13.94	13
138	2159	13883	1.86	11.95	11
137	2808	11724	2.42	10.09	9
136	1990	8916	1.71	7.67	7
135	1522	6926	1.31	5.96	5
134	1293	5404	1.11	4.65	4

133	1108	4111	0.95	3.54	3
132	890	3003	0.77	2.58	2
131	573	2113	0.49	1.82	2
130	566	1540	0.49	1.33	1
129	413	974	0.36	0.84	1
128	193	561	0.17	0.48	1
127	152	368	0.13	0.32	1
126	94	216	0.08	0.19	1
125	54	122	0.05	0.10	1
124	25	68	0.02	0.06	1
123	11	43	0.01	0.04	1
Less Than 123	32	32	0.03	0.03	1

State North Caroline End-Of-Course Testing Program
Geometry—2007

Summary Statistics on Scale Scores

Number of Students With Valid Scores	80300	High Score	179
		Low Score	123
Mean	151.1	State Percentiles	Scale Score
		90	163.8
Standard Deviation	9.6	75	157.9
		50 (Median)	151.1
Variance	92.4	25	143.9
		10	138.4

Frequency Distribution

Scale Score	Frequency	Cumulative Frequency	Percent	Cumulative Percent	2006-07 State Percentiles
179	39	80300	0.05	100.00	99
178	75	80261	0.09	99.95	99
177	50	80186	0.06	99.86	99
176	153	80136	0.19	99.80	99
175	175	79983	0.22	99.61	99
174	108	79808	0.13	99.39	99
173	262	79700	0.33	99.25	99
172	473	79438	0.59	98.93	99
171	380	78965	0.47	98.34	98
170	601	78585	0.75	97.86	97
169	688	77984	0.86	97.12	97
168	761	77296	0.95	96.26	96
167	771	76535	0.96	95.31	95
166	872	75764	1.09	94.35	94
165	1517	74892	1.89	93.27	92
164	1591	73375	1.98	91.38	90
163	1712	71784	2.13	89.39	88
162	1835	70072	2.29	87.26	86
161	1494	68237	1.86	84.98	84
160	2779	66743	3.46	83.12	81
159	2099	63964	2.61	79.66	78
158	2648	61865	3.30	77.04	75
157	2773	59217	3.45	73.74	72
156	2819	56444	3.51	70.29	69
155	2879	53625	3.59	66.78	65
154	3067	50746	3.82	63.20	61
153	2485	47679	3.09	59.38	58
152	4176	45194	5.20	56.28	54
151	2105	41018	2.62	51.08	50
150	2671	38913	3.33	48.46	47
149	3873	36242	4.82	45.13	43
148	2693	32369	3.35	40.31	39
147	2195	29676	2.73	36.96	36
146	3339	27481	4.16	34.22	32
145	2224	24142	2.77	30.06	29
144	3343	21918	4.16	27.30	25
143	2214	18575	2.76	23.13	22
142	2154	16361	2.68	20.37	19
141	2111	14207	2.63	17.69	16
140	2016	12096	2.51	15.06	14
139	1889	10080	2.35	12.55	11
138	1824	8191	2.27	10.20	9
137	854	6367	1.06	7.93	7
136	1108	5513	1.38	6.87	6
135	1297	4405	1.62	5.49	5
134	1032	3108	1.29	3.87	3
133	773	2076	0.96	2.59	2
132	638	1303	0.79	1.62	1
131	305	665	0.38	0.83	1
130	220	360	0.27	0.45	1
129	98	140	0.12	0.17	1
128	23	42	0.03	0.05	1
Less Than 128	19	19	0.02	0.02	1

State North Caroline End-Of-Course Testing Program
Geometry—2007

Summary Statistics on Scale Scores

Number of Students With Valid Scores	80300	High Score	179
		Low Score	123
Mean	151.1	State Percentiles	Scale Score
		90	163.8
Standard Deviation	9.6	75	157.9
		50 (Median)	151.1
Variance	92.4	25	143.9
		10	138.4

Frequency Distribution

Scale Score	Frequency	Cumulative Frequency	Percent	Cumulative Percent	2006–07 State Percentiles
179	39	80300	0.05	100.00	99
178	75	80261	0.09	99.95	99
177	50	80186	0.06	99.86	99
176	153	80136	0.19	99.80	99
175	175	79983	0.22	99.61	99
174	108	79808	0.13	99.39	99
173	262	79700	0.33	99.25	99
172	473	79438	0.59	98.93	99
171	380	78965	0.47	98.34	98
170	601	78585	0.75	97.86	97
169	688	77984	0.86	97.12	97
168	761	77296	0.95	96.26	96
167	771	76535	0.96	95.31	95
166	872	75764	1.09	94.35	94
165	1517	74892	1.89	93.27	92
164	1591	73375	1.98	91.38	90
163	1712	71784	2.13	89.39	88
162	1835	70072	2.29	87.26	86
161	1494	68237	1.86	84.98	84
160	2779	66743	3.46	83.12	81
159	2099	63964	2.61	79.66	78
158	2648	61865	3.30	77.04	75
157	2773	59217	3.45	73.74	72
156	2819	56444	3.51	70.29	69
155	2879	53625	3.59	66.78	65
154	3067	50746	3.82	63.20	61
153	2485	47679	3.09	59.38	58
152	4176	45194	5.20	56.28	54
151	2105	41018	2.62	51.08	50
150	2671	38913	3.33	48.46	47
149	3873	36242	4.82	45.13	43
148	2693	32369	3.35	40.31	39
147	2195	29676	2.73	36.96	36
146	3339	27481	4.16	34.22	32
145	2224	24142	2.77	30.06	29
144	3343	21918	4.16	27.30	25
143	2214	18575	2.76	23.13	22
142	2154	16361	2.68	20.37	19
141	2111	14207	2.63	17.69	16
140	2016	12096	2.51	15.06	14
139	1889	10080	2.35	12.55	11
138	1824	8191	2.27	10.20	9
137	854	6367	1.06	7.93	7
136	1108	5513	1.38	6.87	6
135	1297	4405	1.62	5.49	5
134	1032	3108	1.29	3.87	3
133	773	2076	0.96	2.59	2
132	638	1303	0.79	1.62	1
131	305	665	0.38	0.83	1
130	220	360	0.27	0.45	1
129	98	140	0.12	0.17	1
128	23	42	0.03	0.05	1
Less Than 128	19	19	0.02	0.02	1

Appendix F – *Testing Code of Ethics*

Testing Code of Ethics

Introduction

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

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

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

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

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

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

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

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

.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 G – Excerpts from the Draft Alignment Report

The following pages are excerpts from the noncitable draft of the Alignment Report for the End-of-Grade Tests of Mathematics in grades 3 through 8.

DRAFT for Review

REPORT

Alignment Analysis of Mathematics Standards and Assessments

North Carolina
Grades 3–8 and Algebra

Norman L. Webb

April 30, 2006

This study is one of four alignment studies that will be conducted for the State of North Carolina. An Alignment Analysis Institute was held February 22–24, 2005, in Raleigh, North Carolina, to analyze the mathematics standards and curriculum for grades 3–8, algebra, and alternate assessments. The report consists of a description of the four criteria used to judge the alignment between North Carolina Academic Content Standards for mathematics and multiple assessment forms for grades 3 through 9. This report includes tables listing the results of six reviewers' coding of the assessments and standards.

Acknowledgements

External Reviewers:

Grades 3–5

Michael Kestner, Group Leader	VA
Mary Lynn Raith	PA
Kristen Kalymon	WI

Grades 6 Algebra

Charles Wayne, Group Leader	PA
Linda Wilson	DC
Brian Vesperman	WI

North Carolina Reviewers:

Grades 3–5
Karen Ellis
Timonthy Daly
Michelle Gray

Grades 6 Algebra
Velvet Simington
Judy Musgrave
Carolyn Davis

The State of North Carolina funded this analysis. Dr. Mildred Bazemore, Project Coordinator, and Dr. Nadine McBride, North Carolina Department of Public Instruction, were the main contact people for the Department and oversaw the coordination of the study.

Table of Contents

Executive Summary

Two groups of reviewers of six reviewers each analyzed the alignment among the grades 3 through algebra assessments and the competency goal at a three-day institute held in Raleigh, North Carolina, on February 22–24, 2006. Half of the reviewers were from North Carolina and half were from other states. One group of reviewers analyzed the assessments and competency goals for grades 3–5 and one group analyzed the assessments and competency goals for grades 6 through algebra. Two or three assessment forms were analyzed for each grade or course.

The alignment between the assessments and the grade-level competency goals varied by grade. In general, the analyses indicated that the assessments and competency goals were reasonably or fully aligned for grades 6 and 8. There only needs to be slight improvement on the assessments or competency goals to achieve full alignment for grades 3, 7, and algebra. A greater number of assessment items need to be replaced or modified to achieve full alignment for grades 4 and 5. Over all grades and assessment forms, there were an adequate number of items measuring content related to each competency with an adequate range in coverage and balance. The only exception was that reviewers found only one or two items that related to competency goal 7.2 (Measurement) on the grade 7 Form 124.

The main alignment issue was that the assessments and the competency goals did not meet an acceptable level on the Depth-of-Knowledge Consistency criterion. That is, too many of the assessment items had a depth-of-knowledge (DOK) level that was lower than the level of the corresponding objective. Even though reviewers judged that most objectives (60 to 100 percent) had a DOK level of 1 (Recall and recognition) or 2 (Skill and concepts), too large a proportion of the items were judged to be at DOK Level 1. For example, reviewers judged that 11% of the grade 3 objectives had a DOK level of 1, but they found 40% of the items on a grade 3 assessment form to have a DOK level of 1. Replacing items with more complex items could improve nearly all of the alignment issues. For grades 3, 7, 8, and algebra, only from one to six items would need to be replaced to achieve full alignment. For grades 4 and 5, from seven to 12 items would need to be replaced to achieve full alignment. Reviewers' comments supported the need for a higher proportion of items at DOK Level 2 or 3.

Two or more reviewers coded eight items from grade 8 Form 124 and four items from grade 8 Form 134 to generic objectives. This indicates that reviewers did not find a grade objective that matched these items, which may represent a problem in the wording of the grade 8 objectives in that the statements do not cover all of the content intended to be assessed.

Reviewers had reasonably high agreement in assigning a DOK level of items (generally .80 and higher) and in assigning items to competency goals (.80 and higher) and objectives (.60 and higher).

Alignment Analysis of Mathematics Standards and Assessments

North Carolina
Grades 3–9

Norman L. Webb

Introduction

The alignment of expectations for student learning with assessments for measuring students' attainment of these expectations is an essential attribute for an effective standards-based education system. Alignment is defined as the degree to which expectations and assessments are in agreement and serve in conjunction with one another to guide an education system toward students' learning what they are expected to know and do. As such, alignment is a quality of the relationship between expectations and assessments and not an attribute of any one of these two system components. Alignment describes the match between expectations and assessment that can be legitimately improved by changing either student expectations or the assessments. As a relationship between two or more system components, alignment is determined by using the multiple criteria described in detail in a National Institute for Science Education (NISE) research monograph, *Criteria for Alignment of Expectations and Assessments in Mathematics and Science Education* (Webb, 1997).

A three-day Alignment Analysis Institute was conducted February 22–24, 2006, in Raleigh, North Carolina. Two groups of six reviewers each, including mathematics content experts, district mathematics supervisors, and mathematics teachers analyzed the agreement between the state's mathematics learning goals and assessments. One group analyzed grades 3–5 and one group analyzed grades 6–9. Each set of standards was compared to the assessment administered in the spring of each year.

For the purposes of this analysis, we have employed the convention of standards and objectives to describe two levels of expectations for what students are to know and do. Standard, as used here, refers to Competency Goal (for instance, the learner will understand and use data and simple probability concepts.). Each standard is comprised of up to 12 objectives, sometimes referred to as skills and strategies. These objectives are intended to span the content of the competency goals (standards) under which they fall. The competency goals and objectives are reproduced in Appendix A.

As part of the alignment institute, reviewers were trained to identify the depth-of-knowledge of the competency goals and assessment items. This training included reviewing the definitions of the four depth-of-knowledge (DOK) levels and then reviewing examples of each. Then the reviewers participated in 1) a consensus process to determine the depth-of-knowledge levels of the competency goals and 2) individual analyses of the assessment items. Following individual analyses of the items, reviewers participated in a debriefing discussion in which they assessed the degree to which they had coded particular items or types of content to the standards. After completing the coding for a grade level, the group leaders were given specific assessment items for which a majority of the reviewers had not agreed on the assigned objective. The group of reviewers discussed these items to determine if they could agree on the appropriate corresponding objective. Even with this discussion, for some items and objectives the wording is such that an item could be justifiably coded to more than one objective. After the discussion, reviewers could change their coding if there was a compelling argument to do so.

To derive the results from the analysis, the reviewers' responses are averaged. Any variance among reviewers is considered legitimate, with the true depth-of-knowledge level for the item falling somewhere between two or more assigned values. Such variation could signify a lack of clarity in how the goals were written, the robustness of an item that can legitimately correspond to more than one goal, and/or a depth of knowledge that falls between two of the four defined levels. Reviewers were allowed to identify one assessment item as corresponding to up to three goals—one primary hit (goal) and up to two secondary hits. However, reviewers could only code one depth-of-knowledge level to each assessment item, even if the item corresponded to more than one goal.

Reviewers were instructed to focus primarily on the alignment between the state competency goals and assessments. However, reviewers were encouraged to offer their opinions on the quality of the goals, or of the assessment activities/items, by writing a note about the item. Reviewers could also indicate whether there was a source-of-challenge issue with the item—i.e., a problem with the item that might cause the student who knows the material to give a wrong answer, or enable someone who does not have the knowledge being tested to answer the item correctly.

The results produced from the institute pertain only to the issue of alignment between the North Carolina state competency goals and the state assessment instruments. Note that this alignment analysis does not serve as external verification of the general quality of the state's competency goals or assessments. Rather, only the degree of alignment is discussed. For these results, the means of the reviewers' coding were used to determine whether the alignment criteria were met. When reviewers did vary in their judgments, the means lessened the error that might result from any one reviewer's finding.

This report describes the results of an alignment study of the competency goals and the operational tests in mathematics for grades 3–9 in North Carolina. The study addressed specific criteria related to the content agreement between the state goals and grade-level assessments. Four criteria received major attention: categorical concurrence, depth-of-knowledge consistency, range-of-knowledge correspondence, and balance of representation.

Alignment Criteria Used for This Analysis

This analysis judged the alignment between the goals and the assessment on the basis of four criteria. Information is also reported on the quality of items by identifying items with sources of challenge and other issues. For each alignment criterion, an acceptable level was defined by what would be required to assure that a student had met the goals.

Categorical Concurrence

An important aspect of alignment between goals and assessments is whether both address the same content categories. The categorical-concurrence criterion provides a very general indication of alignment if both documents incorporate the same content. The criterion of categorical concurrence between goals and assessment is met if the same or consistent categories of content appear in both documents. This criterion was judged by determining whether the assessment included items measuring content from each goal. The analysis assumed that the assessment had to have at least six items measuring content from a goal in order for an acceptable level of categorical concurrence to exist between the goal and the assessment. The number of items, six, is based on estimating the number of items that could produce a reasonably reliable subscale for estimating students' mastery of content on that subscale. Of course, many factors have to be considered in determining what a reasonable number is, including the reliability of the subscale, the mean score, and cutoff score for determining mastery. Using a procedure developed by Subkoviak (1988) and assuming that the cutoff score is the mean and that the reliability of one item is .1, it was estimated that six items would produce an agreement coefficient of at least .63. This indicates that about 63% of the group would be consistently classified as masters or nonmasters if two equivalent test administrations were employed. The agreement coefficient would increase if the cutoff score is increased to one standard deviation from the mean to .77 and, with a cutoff score of 1.5 standard deviations from the mean, to .88. Usually states do not report student results by goals or require students to achieve a specified cutoff score on subscales related to a goal. If a state did do this, then the state would seek a higher agreement coefficient than .63. Six items were assumed as a minimum for an assessment measuring content knowledge related to a goal, and as a basis for making some decisions about students' knowledge of that goal. If the mean for six items is 3 and one standard deviation is one item, then a cutoff score set at 4 would produce an agreement coefficient of .77. Any fewer items with a mean of one-half of the items would require a cutoff that would only allow a student to miss one item. This would be a very stringent requirement, considering a reasonable standard error of measurement on the subscale.

Depth-of-Knowledge Consistency

Goals and assessments can be aligned not only on the category of content covered by each, but also on the basis of the complexity of knowledge required by each. Depth-of-knowledge consistency between goals and assessment indicates alignment if what is elicited from students on the assessment is as demanding cognitively as what students are expected to know and do as stated in the goals. For consistency to exist between the assessment and the goal, as judged in this analysis, at least 50% of the items corresponding to a goal had to be at or above the level of knowledge of the goal: 50%, a conservative cutoff point, is based on the assumption that a minimal passing score for any one goal of 50% or higher would require the student to successfully answer at least some items at or above the depth-of-knowledge level of the corresponding goal. For example, assume an assessment included six items related to one goal and students were required to answer correctly four of those items to be judged proficient—i.e., 67% of the items. If three, 50%, of the six items were at or above the depth-of-knowledge level of the corresponding goals, then for a student to achieve a proficient score would require the student to answer correctly at least one item at or above the depth-of-knowledge level of one goal. Some leeway was used in this analysis on this criterion. If a goal had between 40% and 50% of items at or above the depth-of-knowledge levels of the goals, then it was reported that the criterion was "weakly" met.

Interpreting and assigning depth-of-knowledge levels to both standards within goals and assessment items is an essential requirement of alignment analysis. These descriptions help to clarify what the different levels represent in mathematics:

Level 1 (Recall) includes the recall of information such as a fact, definition, term, or a simple procedure, as well as performing a simple algorithm or applying a formula. That is, in mathematics, a one-step, well-defined, and straight algorithmic procedure should be included at this lowest level. Other key words that signify a Level 1 include "identify," "recall," "recognize," "use," and "measure." Verbs such as "describe" and "explain" could be classified at different levels, depending on what is to be described and explained.

Level 2 (Skill/Concept) includes the engagement of some mental processing beyond a habitual response. A Level 2 assessment item requires students to make some decisions as to how to approach the problem or activity, whereas Level 1 requires students to demonstrate a rote response, perform a well-known algorithm, follow a set procedure (like a recipe), or perform a clearly defined series of steps. Keywords that generally distinguish a Level 2 item include "classify," "organize," "estimate," "make observations," "collect and display data," and "compare data." These actions imply more than one step. For example, to compare data requires first identifying characteristics of the objects or phenomenon and then grouping or ordering the objects. Some action verbs, such as "explain," "describe," or "interpret," could be classified at different levels, depending on the object of the action. For example, interpreting information from a simple graph or requiring mathematics information from the graph, is also at Level 2. Interpreting information from a complex graph that requires some decisions on what features of the graph need to be considered and how information from the graph can be aggregated is at Level 3. Level 2 activities are not limited solely to number skills but can involve visualization skills and probability skills. Other Level 2 activities include noticing and describing nontrivial patterns; explaining the purpose and use of experimental procedures; carrying out experimental procedures; making observations and collecting data; classifying, organizing, and comparing data; and organizing and displaying data in tables, graphs, and charts.

Level 3 (Strategic Thinking) requires reasoning, planning, using evidence, and a higher level of thinking than the previous two levels. In most instances, requiring students to explain their thinking is at Level 3. Activities that require students to make conjectures are also at this level. The cognitive demands at Level 3 are complex and abstract. The complexity does not result from the fact that there are multiple answers, a possibility for both Levels 1 and 2, but because the task requires more demanding reasoning. An activity, however, that has more than one possible answer and requires students to justify the response they give would most likely be at Level 3. Other Level 3 activities include drawing conclusions from observations;

citing evidence and developing a logical argument for concepts; explaining phenomena in terms of concepts; and using concepts to solve problems.

Level 4 (Extended Thinking) requires complex reasoning, planning, development, and thinking most likely over an extended period of time. The extended time period is not a distinguishing factor if the required work is only repetitive and does not require applying significant conceptual understanding and higher-order thinking. For example, if a student has to take the water temperature from a river each day for a month and then construct a graph, this would be classified as a Level 2. However, if the student is to conduct a river study that requires taking into consideration a number of variables, this would be at Level 4. At Level 4, the cognitive demands of the task should be high and the work should be very complex. Students should be required to make several connections—relate ideas *within* the content area or *among* content areas—and have to select one approach among many alternatives on how the situation should be solved, in order to be at this highest level. Level 4 activities include developing and proving conjectures; designing and conducting experiments; making connections between a finding and related concepts and phenomena; combining and synthesizing ideas into new concepts; and critiquing experimental designs.

Range-of-Knowledge Correspondence

For goals and assessments to be aligned, the breadth of knowledge required on both should be comparable. The range-of-knowledge criterion is used to judge whether a comparable span of knowledge expected of students by a goal is the same as, or corresponds to, the span of knowledge that students need in order to correctly answer the assessment items/activities. The criterion for correspondence between span of knowledge for a goal and an assessment considers the number of standards within the goal with one related assessment item/activity. Fifty percent of the standards for a goal had to have at least one related assessment item in order for the alignment on this criterion to be judged acceptable. This level is based on the assumption that students' knowledge should be tested on content from over half of the domain of knowledge for a goal. This assumes that each standard for a goal should be given equal weight. Depending on the balance in the distribution of items and the need to have a low number of items related to any one standard, the requirement that assessment items need to be related to more than 50% of the standards for a goal increases the likelihood that students will have to demonstrate knowledge on more than one standard per goal to achieve a minimal passing score. As with the other criteria, a state may choose to make the acceptable level on this criterion more rigorous by requiring an assessment to include items related to a greater number of the standards. However, any restriction on the number of items included on the test will place an upper limit on the number of standards that can be assessed. Range-of-knowledge correspondence is more difficult to attain if the content expectations are partitioned among a greater number of goals and a large number of standards. If 50% or more of the standards for a goal had a corresponding assessment item, then the range-of-knowledge criterion was met. If between 40% and 50% of the standards for a goal had a corresponding assessment item, the criterion was "weakly" met.

Balance of Representation

In addition to comparable depth and breadth of knowledge, aligned goals and assessments require that knowledge be distributed equally in both. The range-of-knowledge criterion only considers the number of standards within a goal hit (a goal with a corresponding item); it does not take into consideration how the hits (or assessment items/activities) are distributed among these standards. The balance-of-representation criterion is used to indicate the degree to which one standard is given more emphasis on the assessment than another. An index is used to judge the distribution of assessment items. This index only considers the standards for a goal that have at least one hit—i.e., one related assessment item per standard. The index is computed by considering the difference in the proportion of standards and the proportion of hits assigned to the standard. An index value of 1 signifies perfect balance and is obtained if the hits (corresponding items) related to a goal are equally distributed among the standards for the given goal. Index values that approach 0 signify that a large proportion of the hits are on only one or two of all of the standards hit. Depending on the number of standards and the number of hits, a unimodal distribution (most items related to one standard and only one item related to each of the remaining standards) has an index value of less than .5. A bimodal distribution has an index value of around .55 or .6. Index values of .7 or higher indicate that items/activities are distributed among all of the standards at least to some degree (e.g., every standard has at least two items) and is used as the acceptable level on this criterion. Index values between .6 and .7 indicate the balance-of-representation criterion has only been "weakly" met.

Source-of-Challenge Criterion

The source-of-challenge criterion is only used to identify items on which the major cognitive demand is inadvertently placed and is other than the targeted mathematics skill, concept, or application. Cultural bias or specialized knowledge could be reasons for an item to have a source-of-challenge problem. Such item characteristics may result in some students not answering an assessment item, or answering an assessment item incorrectly or at a lower level, even though they possess the understanding and skills being assessed.

Findings

Standards

Nine reviewers participated in the depth-of-knowledge (DOK) level consensus process for the competency goals and objectives for the North Carolina Mathematics standards. The six reviewers in each grade-level group (3–5 and 6–9) were joined by three reviewers from the alternate assessment group to determine the DOK levels of the objectives. A summary of their deliberations is presented in Table 1. The complete group consensus values for each competency goal and objective can be found in Appendix A.

Table 1
Percent of Objectives by Depth-of-Knowledge (DOK) Levels for Grades 3–8 and Algebra Standards, North Carolina Alignment Analysis for Mathematics

Grade	Total number of objectives	DOK Level	# of objectives by Level	% within standard by Level
3	17	1 2 3	2 13 2	11 76 11
4	17	1 2 3 4	1 10 5 1	5 58 29 5
5	15	1 2 3	1 8 6	6 53 40
6	23	1 2 3	6 15 2	26 65 8
7	17	1 2 3	1 13 3	5 76 17
8	17	1 2	2 12	14 85
9 Algebra	21	1 2 3	5 15 1	23 71 4

Across the grades, reviewers judged that the majority of objectives had a DOK level of 1 (Recall and Recognition), or a DOK level of 2 (Skills and Concepts). Reviewers judged that a higher proportion of grade 5 objectives were at DOK Level 3 (40%) than was the case for any other grade. For the other grades, reviewers judged that less than 35% of the objectives were at DOK Level 3 or higher. These results indicate that the mathematics expectations generally require students to have _______________ skills and a conceptual understanding of mathematics.

The reviewers were told that within each competency goal the objectives were intended to fully span the content of that competency goal, and, in turn, each competency goal is spanned by the objectives that fall under it. For this reason, the reviewers only coded items to a competency goal if there were no objectives that the item appeared to target. Except for grade 8, reviewers were able to identify a specific objective that matched each item. There were only a few isolated items (three or fewer) on the grades 5, 6, 7, and 9 assessments that at

Table 2

Items Coded to Generic Objectives by More Than One Reviewer, North Carolina Alignment Analysis for Mathematics, Grades 3–9

Grade	Assessment Item	Generic Objective (Number of Reviewers)
3, Form 125	None	
3, Form 234	None	
3, Form 235	None	
4, Form 134	None	
4, Form 234	None	
4, Form 235	None	
5, Form 124	28	5.2 (2)
5, Form 124	24	5.3 (5)
5, Form 125	26	5.3 (2)
5, Form 235	None	
6, Form 124	None	
6, Form 234	35	6.5 (6)
6, Form 235	None	
7, Form 124	None	
7, Form 134	17	7.2 (2)
7, Form 235	None	
8, Form 124	4	8.2 (4)
8, Form 124	6	8.2 (5)
8, Form 124	48	8.2 (4)
8, Form 124	28	8.3 (6)
8, Form 124	10	8.4 (4)
8, Form 124	12	8.4 (5)
8, Form 124	30	8.4 (3)
8, Form 124	51	8.4 (3)
8, Form 134	8	8.2 (5)
8, Form 134	13	8.4 (2)
8, Form 134	31	8.4 (2)
8, Form 134	34	8.5 (2)
9, Algebra IA	None	
9, Algebra IB	None	
9, Algebra IC	26	A3 (4)

least two reviewers coded to a generic objective. However, on the grade 8 assessments, two or more reviewers coded 12 items to a generic objective, 8 items from Form 124 and 4 items from Form 134. These results indicate that reviewers could not find an objective that included the mathematics that was measured by the assessment items. The reviewers' notes reveal the particular issues that reviewers found. For example, grade 8 Form 124 Item 4 required the use of proportions or proportional reasoning, but four of the reviewers did not find any mention of proportions in the grade 8 objectives. Item 48 on the same form required some understanding of volume, but four reviewers did not find an appropriate grade 8 objective, but report the item corresponded better to a grade 7 objective. The high number of grade 8 items on Form 124 assigned to generic objectives should be reviewed to determine whether the reviewers' judgments are valid. The alignment issue may have more to do with an omission of a mathematical idea from the standards than with the assessment. For the other grades, the absence of any generic objectives indicates that the set of standards is written with adequate clarity and covers the tested domain such that reviewers were able to find an objective that matched each item on each form.

Alignment of Curriculum Standards and Assessments

The assessments for grades 3–7 were each comprised of 50 multiple-choice items. The grade 8 assessments had 60 multiple-choice items, and the algebra assessments had 80 multiple-choice items. All of the items were worth one point.

The results of the analysis for each of the four alignment criteria are summarized in Table 3. In Table 3, "YES" indicates that an acceptable level was attained between the assessment and the standard on the criterion. "WEAK" indicates that the criterion was nearly met, within a margin that could simply be due to error in the system. "NO" indicates that the criterion was not met by a noticeable margin—10% over an acceptable level for Depth-of-Knowledge Consistency, 10% over an acceptable level for Range-of-Knowledge Correspondence, and .1 under an index value of .7 for Balance of Representation.

Overall, the alignment between the competency goals and the assessment needs slight improvement. Across all of the grades and forms, from zero to 12 items would need to be modified or replaced to achieve full alignment. Reviewers found the grade 6 mathematics assessments and competency goals to be fully aligned. No changes at grade 6 are necessary. The main alignment issue at the other grades is that the depth-of-knowledge levels of the items are too low in comparison to the depth-of-knowledge levels of the objectives. The alignment for each grade is discussed in more detail below.

Grade 3

Each of the three grade 3 forms analyzed had a sufficient number for each competency goal that assessed at least some content for an adequate number of objectives under each goal and were adequately distributed among these objectives. However, an acceptable level on the Depth-of-Knowledge Consistency criterion was not met on grade 3 Form 125 by four of the five competency goals, on Form 234 by three of the five competency goals, and on Form 235 by two of the five competency goals. For each competency goal and for each form, only one or two items need to be replaced by an item with a higher DOK level to achieve full alignment. The following number of items by competency goal would need to be replaced by more complex items:

Form 125	CG 3.1	2 items
	CG 3.2	2 items

	CG 3.3	1 item
	CG 3.5	1 item
Form 234	CG 3.1	1 item
	CG 3.2	1 item
	CG 3.3	2 items
Form 235	CG 3.2	1 item
	CG 3.3	1 item

Grade 4

As for grade 3, each of the three grade 4 forms analyzed had a sufficient number for each competency goal that assessed at least some content for an adequate number of objectives under each goal and was adequately distributed among these objectives. However, an acceptable level on the Depth-of-Knowledge Consistency criterion was not met for each of the three grade 4 forms by the same three out of five competency goals. For each competency goal and for each form, one to six items need to be replaced by an item with a higher DOK level to achieve full alignment. The following number of items by competency goal would need to be replaced by more complex items:

Form 124	CG 4.1	4 items
	CG 4.2	1 item
	CG 4.4	4 items
Form 234	CG 4.1	3 items
	CG 4.2	2 items
	CG 4.4	3 items
Form 234	CG 4.1	6 items
	CG 4.2	1 item
	CG 4.4	3 items

Table 3
Summary of Acceptable Levels on Alignment Criteria for Mathematics Grades 3–Algebra Assessments for North Carolina Alignment Analysis

Grade 3, Form 125	***Alignment Criteria***			
Standards	*Categorical Concurrence*	*Depth-of-Knowledge Consistency*	*Range of Knowledge*	*Balance of Represent-ation*
3.1–Model, identify, and compute with whole numbers	YES	WEAK	YES	YES
3.2–Metric and customary measurement	YES	NO	YES	YES
3.3–Basic geometric properties of 2- and 3-dimensional figures	YES	NO	YES	YES
3.4–Data and simple probability concepts	YES	YES	YES	YES
3.5–Patterns and simple mathematical relationships	YES	WEAK	YES	YES
Grade 3, Form 234	***Alignment Criteria***			
Standards	*Categorical Concurrence*	*Depth-of-Knowledge Consistency*	*Range of Knowledge*	*Balance of Represent-ation*
3.1–Model, identify, and compute with whole numbers	YES	WEAK	YES	YES
3.2–Metric and customary measurement	YES	NO	YES	YES
3.3–Basic geometric properties of 2- and 3-dimensional figures	YES	NO	YES	YES
3.4–Data and simple probability concepts	YES	YES	YES	YES
3.5–Patterns and simple mathematical relationships	YES	YES	YES	YES
Grade 3, Form 235	***Alignment Criteria***			
Standards	*Categorical Concurrence*	*Depth-of-Knowledge Consistency*	*Range of Knowledge*	*Balance of Represent-ation*
3.1–Model, identify, and compute with whole numbers	YES	YES	YES	YES
3.2–Metric and customary measurement	YES	WEAK	YES	YES
3.3–Basic geometric properties of 2- and 3-dimensional figures	YES	NO	YES	YES
3.4–Data and simple probability concepts	YES	YES	YES	YES
3.5–Patterns and simple mathematical relationships	YES	YES	YES	YES

Table 3 (continued)
Summary of Acceptable Levels on Alignment Criteria for Mathematics Grades 3-Algebra Assessments for North Carolina Alignment Analysis

Grade 4, Form 134	***Alignment Criteria***			
Standards	*Categorical Concurrence*	*Depth-of-Knowledge Consistency*	*Range of Knowledge*	*Balance of Represent-ation*
4.1–Non-negative rational numbers	YES	NO	YES	YES
4.2–Perimeter and area	YES	NO	YES	YES
4.3–Geometric properties and relationships	YES	YES	YES	YES
4.4–Graphs, probability, and data analysis	YES	NO	YES	YES
4.5–Mathematical relationships	YES	YES	YES	YES
Grade 4, Form 234	***Alignment Criteria***			
Standards	*Categorical Concurrence*	*Depth-of-Knowledge Consistency*	*Range of Knowledge*	*Balance of Represent-ation*
4.1–Non-negative rational numbers	YES	NO	YES	YES
4.2–Perimeter and area	YES	NO	YES	YES
4.3–Geometric properties and relationships	YES	YES	YES	YES
4.4–Graphs, probability, and data analysis	YES	NO	YES	YES
4.5–Mathematical relationships	YES	YES	YES	YES
Grade 4, Form 235	***Alignment Criteria***			
Standards	*Categorical Concurrence*	*Depth-of-Knowledge Consistency*	*Range of Knowledge*	*Balance of Represent-ation*
4.1–Non-negative rational numbers	YES	NO	YES	YES
4.2–Perimeter and area	YES	NO	YES	YES
4.3–Geometric properties and relationships	YES	YES	YES	YES
4.4–Graphs, probability, and data analysis	YES	NO	YES	YES
4.5–Mathematical relationships	YES	YES	YES	YES

Grade 5

Reviewers did not find a sufficient number of items, at least six, for two competency goals, 5.2 and 5.4, on grade 5 Form 124. For each of these competency goals, one more item needs to be added. To achieve an acceptable level on the Depth-of-Knowledge Consistency criterion for each of the five competency goals on each form would require from one to four items to be replaced or modified for each competency goal. The following number of items by competency goal would need to be replaced by items that are more complex or added:

Form 124	CG 5.1	3 items
	CG 5.2	1 item added with appropriate DOK level
	CG 5.4	1 item added with appropriate DOK level
	CG 5.5	2 items
Form 125	CG 5.1	4 items
	CG 5.2	1 item
	CG 5.3	2 items
	CG 5.4	3 items
	CG 5.5	2 items
Form 235	CG 5.1	3 items
	CG 5.2	1 item
	CG 5.4	3 items
	CG 5.5	4 items

Grade 6

The three grade 6 assessment forms and the competency goals were found to be fully aligned.

Grade 7

Reviewers found only one item on grade 7 Form 124 that corresponded to competency goal 7.2 (Measurement). Five items on this form need to be replaced by measurement items corresponding to competency goal 7.2 to achieve full alignment. All of the other alignment criteria were fully met for grade 7 Form 124 and the standards. The other two grade 7 forms, 134 and 235, had sufficient items for each of the five standards with appropriate range and balance but did not fully meet the Depth-of-Knowledge Consistency criterion. One or two items need to be replaced for three of the five competencies on each of these forms to achieve full alignment. The following number of items by competency goal would need to be replaced by items that are more complex or added:

Form 124	CG 7.2	5 items (added)
Form 134	CG 7.1	2 items
	CG 7.4	1 item
	CG 7.5	1 item
Form 235	CG 7.1	1 item
	CG 7.4	2 items
	CG 7.5	1 item

Table 3 (continued)
Summary of Acceptable Levels on Alignment Criteria for Mathematics Grades 3–Algebra Assessments for North Carolina Alignment Analysis

Grade 5, Form 124	***Alignment Criteria***			
Standards	*Categorical Concurrence*	*Depth-of-Knowledge Consistency*	*Range of Knowledge*	*Balance of Represent-ation*
5.1–Non-negative rational numbers	YES	NO	YES	YES
5.2–Metric and customary measurement	NO	NO	YES	YES
5.3–Properties and relationships of plane figures	YES	YES	YES	YES
5.4–Graphs and data analysis	NO	NO	YES	YES
5.5–Patterns, relationships, and elementary algebraic representation	YES	NO	YES	YES
Grade 5, Form 125	***Alignment Criteria***			
Standards	*Categorical Concurrence*	*Depth-of-Knowledge Consistency*	*Range of Knowledge*	*Balance of Represent-ation*
5.1–Non-negative rational numbers	YES	NO	YES	YES
5.2–Metric and customary measurement	YES	WEAK	YES	YES
5.3–Properties and relationships of plane figures	YES	WEAK	YES	WEAK
5.4–Graphs and data analysis	YES	NO	YES	YES
5.5–Patterns, relationships, and elementary algebraic representation	YES	NO	YES	YES
Grade 5, Form 235	***Alignment Criteria***			
Standards	*Categorical Concurrence*	*Depth-of-Knowledge Consistency*	*Range of Knowledge*	*Balance of Represent-ation*
5.1–Non-negative rational numbers	YES	NO	YES	YES
5.2–Metric and customary measurement	YES	NO	YES	YES
5.3–Properties and relationships of plane figures	YES	YES	YES	YES
5.4–Graphs and data analysis	YES	NO	YES	YES
5.5–Patterns, relationships, and elementary algebraic representation	YES	NO	YES	YES

Table 3 (continued)
Summary of Acceptable Levels on Alignment Criteria for Mathematics Grades 3-Algebra Assessments for North Carolina Alignment Analysis

Grade 6, Form 124	***Alignment Criteria***			
Standards	*Categorical Concurrence*	*Depth-of-Knowledge Consistency*	*Range of Knowledge*	*Balance of Represent-ation*
6.1–Rational numbers	YES	YES	YES	YES
6.2–Measure two- and three-dimensional figures	YES	YES	YES	YES
6.3–Geometric figures in the coordinate plane	YES	YES	YES	YES
6.4–Probabilities	YES	YES	YES	YES
6.5–Simple algebraic expressions	YES	YES	YES	YES
Grade 6, Form 234	***Alignment Criteria***			
Standards	*Categorical Concurrence*	*Depth-of-Knowledge Consistency*	*Range of Knowledge*	*Balance of Represent-ation*
6.1–Rational numbers	YES	YES	YES	YES
6.2–Measure two- and three-dimensional figures	YES	YES	YES	YES
6.3–Geometric figures in the coordinate plane	YES	YES	YES	YES
6.4–Probabilities	YES	YES	YES	YES
6.5–Simple algebraic expressions	YES	YES	YES	YES
Grade 6, Form 235	***Alignment Criteria***			
Standards	*Categorical Concurrence*	*Depth-of-Knowledge Consistency*	*Range of Knowledge*	*Balance of Represent-ation*
6.1–Rational numbers	YES	YES	YES	YES
6.2–Measure two- and three-dimensional figures	YES	YES	YES	YES
6.3–Geometric figures in the coordinate plane	YES	YES	YES	YES
6.4–Probabilities	YES	YES	YES	YES
6.5–Simple algebraic expressions	YES	YES	YES	YES

Table 3 (continued)
Summary of Acceptable Levels on Alignment Criteria for Mathematics Grades 3-Algebra Assessments for North Carolina Alignment Analysis

Grade 7, Form 124	***Alignment Criteria***			
Standards	*Categorical Concurrence*	*Depth-of-Knowledge Consistency*	*Range of Knowledge*	*Balance of Represent-ation*
7.1–Rational numbers	YES	YES	YES	YES
7.2–Measurement involving two- and three-dimensional figures	NO	NA	NA	NA
7.3–Properties and relationships in geometry	YES	YES	YES	YES
7.4–Graphs and data analysis	YES	YES	YES	YES
7.5–Linear relations and fundamental algebraic concepts	YES	YES	YES	YES
Grade 7, Form 134	***Alignment Criteria***			
Standards	*Categorical Concurrence*	*Depth-of-Knowledge Consistency*	*Range of Knowledge*	*Balance of Represent-ation*
7.1–Rational numbers	YES	NO	YES	YES
7.2–Measurement involving two- and three-dimensional figures	YES	YES	YES	YES
7.3–Properties and relationships in geometry	YES	YES	YES	YES
7.4–Graphs and data analysis	YES	WEAK	YES	YES
7.5–Linear relations and fundamental algebraic concepts	YES	WEAK	YES	YES
Grade 7, Form 235	***Alignment Criteria***			
Standards	*Categorical Concurrence*	*Depth-of-Knowledge Consistency*	*Range of Knowledge*	*Balance of Represent-ation*
7.1–Rational numbers	YES	NO	YES	YES
7.2–Measurement involving two- and three-dimensional figures	YES	YES	YES	YES
7.3–Properties and relationships in geometry	YES	YES	YES	YES
7.4–Graphs and data analysis	YES	NO	YES	YES
7.5–Linear relations and fundamental algebraic concepts	YES	WEAK	YES	YES

Grade 8

The two grade 8 forms and the competency goals were found to be reasonably aligned. Only one or two items on each of the forms need to be replaced by an item with a higher DOK level for full alignment to exist between the assessments and the standards. The following number of items by competency goal would need to be replaced with items that are more complex:

Form 124	CG 8.1	1 item
Form 134	CG 8.1	2 items

Table 3 (continued)
Summary of Acceptable Levels on Alignment Criteria for Mathematics Grades 3–Algebra Assessments for North Carolina Alignment Analysis

Grade 8, Form 124	***Alignment Criteria***			
Standards	*Categorical Concurrence*	*Depth-of-Knowledge Consistency*	*Range of Knowledge*	*Balance of Represent-ation*
8.1– Real numbers	YES	WEAK	YES	YES
8.2–Measurement concepts	YES	YES	YES	YES
8.3–Properties and relationships in geometry	YES	YES	YES	YES
8.4–Graphs and data analysis	YES	YES	YES	YES
8.5–Linear relations and functions	YES	YES	YES	YES

Grade 8, Form 134	***Alignment Criteria***			
Standards	*Categorical Concurrence*	*Depth-of-Knowledge Consistency*	*Range of Knowledge*	*Balance of Represent-ation*
8.1–Real numbers	YES	NO	YES	YES
8.2–Measurement concepts	YES	YES	YES	YES
8.3–Properties and relationships in geometry	YES	YES	YES	YES
8.4–Graphs and data analysis	YES	YES	YES	YES
8.5–Linear relations and functions	YES	YES	YES	YES

[…]

Table 6
Average Reviewer Opinion on Overall Alignment of Assessments for Grades 3–9, North Carolina Mathematics Curriculum Standards for Grades 3-9 (Question D)
(1-Perfect alignment, 2-Acceptable alignment, 3-Needs slight improvement, 4-Needs major improvement, 5-Not aligned in any way)

Assessment	Avg. Response	Number of Reviewers
Grade 3 Form 125	2.5	6
Grade 3 Form 234	2.17	6
Grade 3 Form 235	2.5	6
Grade 4 Form 134	2.6	5
Grade 4 Form 234	2.4	5
Grade 4 Form 235	2.6	5
Grade 5 Form 124	2.4	5
Grade 5 Form 125	2.4	5
Grade 5 Form 235	2.4	5
Grade 6 Form 124	2.5	6
Grade 6 Form 234	2.5	6
Grade 6 Form 235	2.5	6
Grade 7 Form 124	2.0	6
Grade 7 Form 134	2.0	2
Grade 7 Form 235	2.0	5
Grade 8 Form 124	3.0	5
Grade 8 Form 134	3.17	6
Grade 9 Algebra IA	2.75	4
Grade 9 Algebra IB	2.6	5
Grade 9 Algebra IC	2.0	5

Reliability among Reviewers

The overall intraclass correlation among the mathematics reviewers' assignment of DOK levels to items was good (Table 7). An intraclass correlation value greater than 0.8 generally indicates a high level of agreement among the reviewers. The grades 3–5 group had very high agreement in assigning DOK levels to items, eight of the nine analyses had an intraclass correlation of .80 or higher (Table 7). The agreement among the grades 6–9 group was not as high, but reasonable. For nine of the 11 analyses, the grades 6–9 group had an intraclass correlation of .70 or higher.

A pairwise comparison was used to determine the degree of reliability of reviewer coding at the competency goal level and at the objective level (Table 8). The pairwise agreement values for the competency goals were .80 or higher except for one of the 20 analyses. This indicates that the reviewers had strong agreement in assigning items to a competency goal. Reviewers had less agreement in assigning items to specific objectives, but most of the pairwise agreements were reasonable and above .60. Reviewers agreed less in assigning items to objectives in analyzing grade 7 forms 124 and 134. These test forms were some of the first forms analyzed by the grade 6–9 group and indicate that the reviewers could have received more training. The agreement among reviewers did improve in assigning items to objectives for the other forms and grades.

Table 7

Intraclass Correlation among Reviewers in Assigning Item Depth-of-Knowledge Level for Mathematics

Grade	Intraclass Correlation	Number of Items	Number of Reviewers
3 Form 125	0.86	50	6
3 Form 234	0.83	50	6
3 Form 235	0.88	50	6
4 Form 134	0.89	50	6
4 Form 234	0.86	50	6
4 Form 235	0.86	50	6
5 Form 124	0.86	50	6
5 Form 125	0.84	50	6
5 Form 235	0.74	50	6
6 Form 124	0.71	50	6
6 Form 234	0.87	50	6
6 Form 235	0.78	50	6
7 Form 124	0.77	50	6
7 Form 134	0.62*	50	4
7 Form 235	0.62	50	6
8 Form 124	0.83	60	6
8 Form 134	0.86	60	6
9 Algebra IA	0.75	80	5
9 Algebra IB	0.83	80	5
9 Algebra IC	0.74	80	6

* Pairwise correlation was used because of low variation among values.

Table 8

Pairwise Agreement among Reviewers in Assigning Items to Competency Goals and Objectives for Mathematics

Grade and Form	Pairwise: Competency Goal	Pairwise: Standard
3 Form 125	.91	.80
3 Form 234	.92	.72
3 Form 235	.94	.84
4 Form 134	.87	.66
4 Form 234	.90	.68
4 Form 235	.91	.68
5 Form 124	.89	.70
5 Form 125	.88	.67
5 Form 235	.88	.68
6 Form 124	.94	.72
6 Form 234	.88	.70
6 Form 235	.90	.69
7 Form 124	.78	.43
7 Form 134	.82	.47
7 Form 235	.86	.64
8 Form 124	.83	.62
8 Form 134	.85	.67
9 Algebra IA	.89	.74
9 Algebra IB	.89	.68
9 Algebra IC	.88	.65

Summary

Two groups of reviewers with six reviewers each analyzed the alignment among the grades 3 through algebra assessments and competency goals at a three-day institute held in Raleigh, North Carolina, on February 22–24, 2006. Half of the reviewers were from North Carolina and half were from other states. One group of reviewers analyzed the assessments and competency goals for grades 3–5 and one group analyzed the assessments and competency goals for grades 6 through algebra. Two or three assessment forms were analyzed for each grade or course.

The alignment between the assessments and the grade-level competency goals varied by grade. In general, the analyses indicated that the assessments and competency goals were reasonably or fully aligned for grades 6 and 8. There only needs to be slight improvement on the assessments or competency goals to achieve full alignment for grades 3, 7, and algebra. A greater number of assessment items need to be replaced or modified to achieve full alignment for grades 4 and 5. Over all grades and assessment forms there were an adequate number of items measuring content related to each competency with an adequate range in coverage and balance. The only exception was that reviewers found only one or two items that related to competency goal 7.2 (measurement) on the grade 7 Form 124.

The main alignment issue was that the assessments and the competency goals did not meet an acceptable level on the Depth-of-Knowledge Consistency criterion. That is, too many of the assessment items had a depth-of-knowledge level that was lower than the level of the corresponding objective. Even though reviewers judged that most objectives (60 to 100 percent) had a DOK level of 1 (recall and recognition) or 2 (skill and concepts), too large a proportion of the items were judged to have a DOK level 1. For example, reviewers judged that 11% of the grade 3 objectives had a DOK level 1, but they found 40% of the items on a grade 3 assessment form to have a DOK level 1. Replacing items with more complex items could improve nearly all of the alignment issues. For grades 3, 7, 8, and algebra only from one to six items would need to be replaced to achieve full alignment. For grades 4 and 5, from seven to 12 items would need to be replaced to achieve full alignment. Reviewers' comments supported the need to have a higher proportion of items with a DOK level 2 or 3.

Two or more reviewers coded eight items from grade 8 Form 124 and four items from grade 8 Form 134 to generic objectives. This indicates that reviewers did not find a grade objective that matched these items. Thus, there may be a problem in the wording of the grade 8 objectives such that the statements do not cover all of the content intended to be assessed.

Reviewers had reasonably high agreement in assigning a DOK level of items (generally .80 and higher) and in assigning items to competency goals (.80 and higher) and objectives (.60 and higher).

References

Subkoviak, M. J. (1988). A practitioner's guide to computation and interpretation of reliability indices for mastery tests. *Journal of Educational Measurement, 25*(1), 47–55.

Valencia, S. W., & Wixson, K. K. (2000). Policy-oriented research on literary standards and assessment. In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research: Vol III.* Mahwah, NJ: Lawrence Erlbaum.

Webb, N. L. (1997). *Criteria for alignment of expectations and assessments in mathematics and science education*. Council of Chief State School Officers and National Institute for Science Education Research Monograph No. 6. Madison: University of Wisconsin, Wisconsin Center for Education Research.

Appendix H – EOG Standard Setting

This Appendix contains only the Executive Summary of the Standard Setting Technical Report.

North Carolina End-Of-Grade Tests
Standard Setting for Mathematics

September 11–15, 2006
Raleigh, North Carolina

EXECUTIVE SUMMARY

Introduction

Committees of North Carolina educators were convened on September 11–16, 2006, in order to set standards for mathematics on the North Carolina End-Of-Grade (EOG) Test. Three committees consisting of a total of 62 educators participated in the five-day conference. The item-mapping procedure was applied to set the standards. The outcomes of the conference are described in this summary and more detailed information is provided in the following Standard Setting Technical Report.

Panelist Information

The conference consisted of the following three committees meeting simultaneously: a grades 3 and 4 committee, a grades 5 and 6 committee, and a grades 7 and 8 committee. Of the 62 panelists, 22 were in the grades 3 and 4 committee, 19 were in the grades 5 and 6 committee, and 21 were in the grades 7 and 8 committee. All 62 educators provided voluntary demographic information.

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.

	Gender		Ethnicity			District Size		
Committee	Male	Female	White	Black	Hispanic	Large	Medium	Small
Grades 3 and 4	4.55	95.45	81.82	13.64	4.55	31.82	54.55	13.64
Grades 5 and 6	21.05	78.95	89.47	10.53	0.00	47.37	21.05	31.58
Grades 7 and 8	9.52	90.48	85.71	14.28	0.00	23.81	42.85	33.33

Method and Procedure

The standard-setting conference began on Monday, September 11. The morning of Monday, September 11, was devoted to introductions to the staff, to standard setting, and to the mathematics EOG test. For this stage of the conference, all panelists met together in one large room.

Following lunch, the committees began the process of creating performance level descriptors for each grade. The three committees (grades 3 and 4, grades 5 and 6, and grade 7 and 8) met separately in individual conference rooms. This process required the afternoon of Monday, September 11, and the morning of Tuesday, September 12. The result from creating performance level descriptors was a set of descriptors for each performance level (Level I, II, III and IV) at each grade level.

After lunch on Tuesday, September 12, the committees began the standard-setting process. The standard-setting process consisted of three rounds of judgments. Panelists were divided into small groups of 5 or 6 members that worked together at a table within the conference room.

The item-mapping procedure was the judgmental process used. In this procedure, panelists are 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.

Ordered item books were constructed from the field test forms available from the last operational tests. The number of items in each ordered item book is shown in Table 2. Items were ordered by adjusting the 0.67 response probability (67RP) for chance by replacing the value 0.67 with (c+2)/3. Items were then sorted from least to most difficult. Each ordered item book was accompanied by an item map containing the following:

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

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

Grade	Number of Items
3	92
4	92
5	92
6	92
7	80
8	80

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 pseudochance 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. That is why items in the ordered item book were sorted using the corrected-for-chance response probability of (c + 2) / 3.

The cut score at each achievement level was determined by computing the median item number across panelists at a given grade level. This represents the minimum raw score that an examinee must attain to be classified at the particular level. Cuts are usually computed to be between raw scores. In the final report, all cut scores will be rounded to the next higher point if the decimal value is larger than 4 (e.g., 15.5 would become 16).

At the beginning of Rounds 2 and 3 panelists were provided with results generated from the previous round to inform their decision making. 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. The medians of all panelists were used to describe the content-group cut scores. Finally, panelists were briefed on the results of their Round 3 ratings.

On Friday, September 15, a subcommittee met to review performance level descriptions and cut scores across grade levels. This subcommittee was comprised of four panelists from each of the three original committees. Initially, the subcommittee was presented with the performance level descriptors from each grade. The subcommittee was tasked with reviewing the performance level descriptors for reasonableness and coherence across grade levels.

Next, this subcommittee was presented with the cut scores and the percentage of students that would be classified in each performance level based on the Round 3 final cut scores. The subcommittee was tasked with determining if the cut scores showed a reasonable pattern across performance levels and across grades.

Results

Performance level Descriptors

The subcommittee approved the performance level descriptor documents on Friday with minor modifications. These documents are presented in the following standard-setting report. The subcommittee members stated that they preferred the format from the grade 5 and grade 6 performance level descriptors. Furthermore, the subcommittee endorsed the use of the initial paragraph from the grade-level indicators to preface each table of performance level descriptors.

Cut Scores

Table 3 summarizes the cut scores after the Round 3 final rating. Table 3 shows the percent score and the page number for each performance level at each grade. The percent score was computed as the percent of the items in the item book. 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.

GRADE	Statistics	LEVEL II		LEVEL III		LEVEL IV	
		Percent Score	Page Number	Percent Score	Page Number	Percent Score	Page Number
Grade 3	Mean	17	15	37	34	72	67
	Median	18	17	37	34	72	66
	Min	11	10	27	25	62	57
	Max	20	18	48	44	79	73
Grade 4	Mean	17	16	41	38	77	71
	Median	16	15	41	38	75	69
	Min	10	9	33	30	71	65
	Max	21	19	47	43	86	79
Grade 5	Mean	13	12	36	33	70	64
	Median	14	13	39	36	68	63
	Min	9	8	24	22	53	49
	Max	20	18	49	45	79	73
Grade 6	Mean	14	13	33	30	68	63
	Median	15	14	33	30	68	63
	Min	12	11	23	21	58	53
	Max	17	16	39	36	84	77
Grade 7	Mean	13	11	41	33	73	59
	Median	11	9	43	34	75	60
	Min	9	7	33	26	60	48
	Max	21	17	53	42	88	70
Grade 8	Mean	18	14	46	37	72	58
	Median	19	15	44	35	73	58
	Min	14	11	41	33	69	55
	Max	21	17	51	41	78	62

Table 4 shows the percent of students that would be categorized in each performance level based on the median rating for the Round 3 final ratings. Note that across grades the percentage of students that would be categorized in Level I is less than 1 percent.

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.

Grade	Level I	Level II	Level III	Level IV
Grade 3	0.73	13.29	55.10	30.88
Grade 4	0.43	21.42	51.51	26.64
Grade 5	0.18	18.31	46.10	35.41
Grade 6	0.32	14.33	56.20	29.15
Grade 7	0.10	31.05	44.48	24.37
Grade 8	0.68	31.33	44.81	23.18

The graph of the percent of students that would be categorized in each performance level based on the median rating for the Round 3 final ratings is shown in Figure 1.

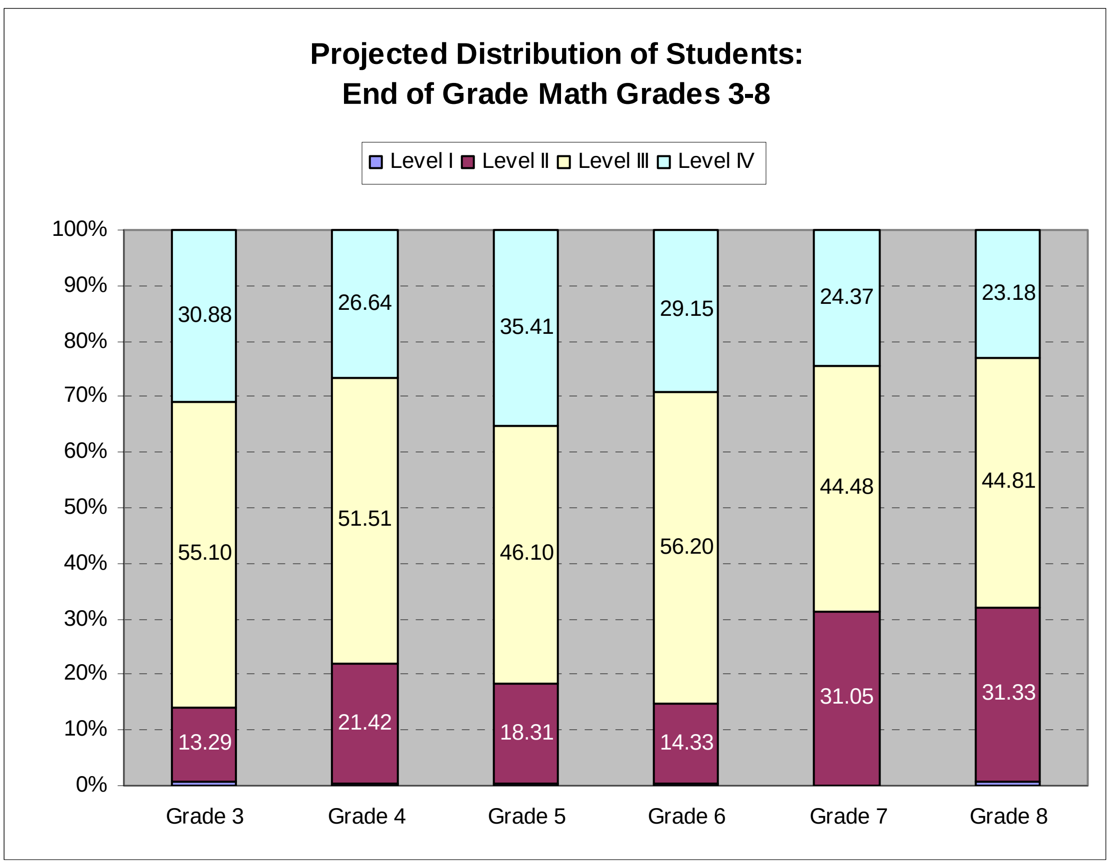


Figure 1. The percent of students that would be categorized in each performance level based on the median rating for the Round 3 final ratings.

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 5.

Table 5. Mean response to questions on the exit survey[1].

	Grade					
Question	3	4	5	6	7	8
The method for setting standards, item mapping, was conceptually clear.	4.73	4.91	4.21	4.50	4.21	4.72
I had a good understanding of what the test was intended to measure.	4.77	4.86	4.74	4.72	4.70	4.67
I could clearly distinguish between student performance levels.	4.41	4.59	4.05	4.22	4.05	4.17
After the first round of ratings, I felt comfortable with the standard-setting procedure.	4.27	4.77	3.53	4.50	4.32	4.56
I found the feedback on item difficulty useful in setting standards.	4.86	4.82	4.47	4.72	4.37	4.61
I found the feedback on the ratings of judges compared to other judges useful in setting standards.	4.68	4.95	4.53	4.61	4.50	4.65
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	4.67	4.20	4.28
I feel confident that the final cut-score recommendations reflect the performance levels associated with the Mathematics EOG Test.	4.32	4.77	4.42	4.72	4.00	4.53

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

Appendix I – EOC Standard Setting

This Appendix contains only the Executive Summary of the Standard Setting Report.

North Carolina End of Course Tests
Standard Setting for Algebra I, Algebra II, Geometry and English I
October 15–16, 2007

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

	Gender		Ethnicity			District Size		
Committee	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 pseudochance 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. Throughout the standard-setting process panelists received correct instructions and valid information from the facilitators. However, they were shown incorrect data on a PowerPoint slide between the determination of their round 2 and round 3 bookmarks. 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

	LEVEL II		LEVEL III		LEVEL IV	
Subject	**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

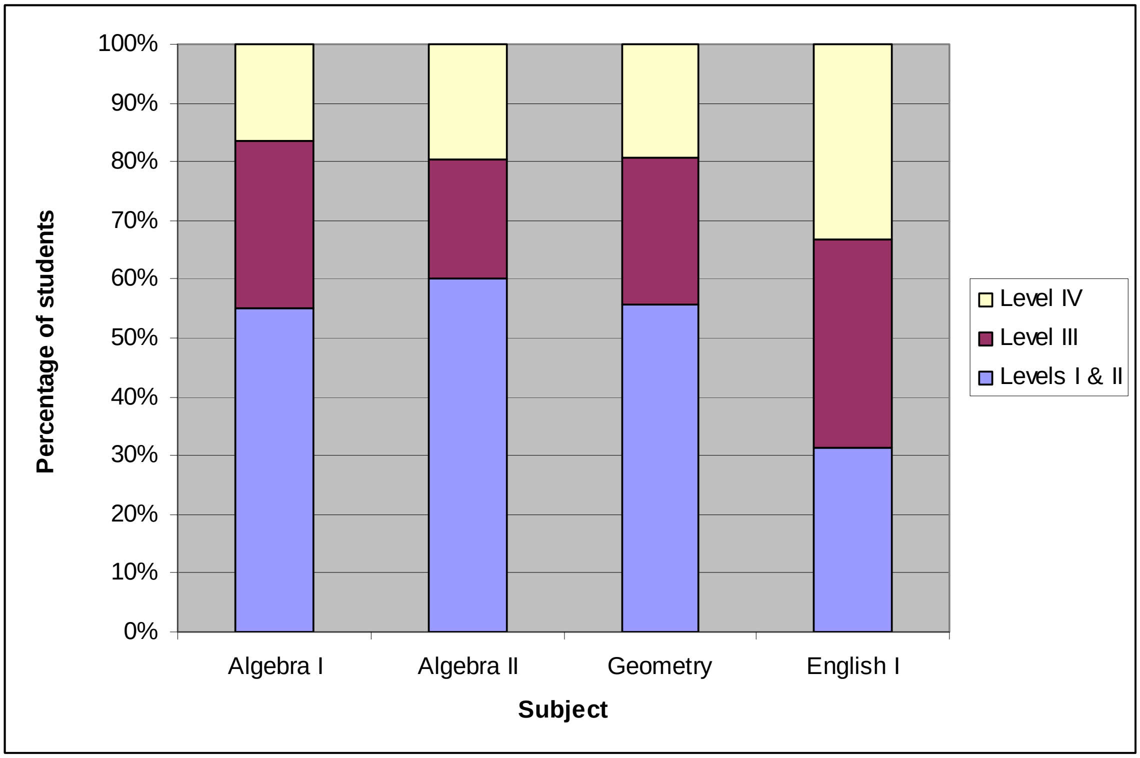
Subject	Levels I & II	Level III	Level IV
Algebra I	55.10	28.30	16.57
Algebra II	59.99	20.41	19.60
Geometry	55.70	25.04	19.26
English I	31.24	35.66	33.10

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 survey[1]

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 first 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 J – Raw to Scale Score Conversion Tables

Grade 3 Pretest Form K

Score	EAP	SD	Score	SE
0	-2.55	0.52	295	7
1	-2.46	0.53	296	7
2	-2.36	0.53	297	7
3	-2.25	0.53	298	7
4	-2.13	0.53	300	7
5	-2.00	0.53	302	7
6	-1.87	0.52	303	7
7	-1.73	0.51	305	6
8	-1.59	0.49	307	6
9	-1.44	0.47	309	6
10	-1.30	0.45	311	6
11	-1.16	0.43	312	5
12	-1.02	0.41	314	5
13	-0.89	0.39	316	5
14	-0.76	0.37	317	5
15	-0.64	0.35	319	4
16	-0.53	0.33	320	4
17	-0.41	0.32	322	4
18	-0.31	0.31	323	4
19	-0.20	0.30	324	4
20	-0.10	0.29	326	4
21	0.00	0.28	327	4
22	0.10	0.28	328	4
23	0.20	0.27	330	3
24	0.30	0.27	331	3
25	0.40	0.27	332	3
26	0.50	0.27	333	3
27	0.60	0.27	335	3
28	0.71	0.27	336	3
29	0.82	0.27	337	3
30	0.93	0.28	339	4
31	1.05	0.28	340	4
32	1.18	0.29	342	4
33	1.32	0.31	344	4
34	1.47	0.33	346	4
35	1.65	0.35	348	4
36	1.86	0.38	351	5
37	2.12	0.43	354	5
38	2.45	0.50	358	6

Grade 3 Pretest form L

Score	EAP	SD	Score	SE
0	-2.45	0.52	296	7
1	-2.37	0.53	297	7
2	-2.29	0.53	298	7
3	-2.19	0.53	299	7
4	-2.08	0.53	301	7
5	-1.97	0.53	302	7
6	-1.84	0.53	304	7
7	-1.71	0.52	305	7
8	-1.57	0.50	307	6
9	-1.43	0.49	309	6
10	-1.29	0.46	311	6
11	-1.15	0.44	312	6
12	-1.01	0.42	314	5
13	-0.88	0.39	316	5
14	-0.75	0.37	317	5
15	-0.63	0.35	319	4
16	-0.52	0.33	320	4
17	-0.40	0.32	322	4
18	-0.30	0.31	323	4
19	-0.19	0.30	325	4
20	-0.09	0.29	326	4
21	0.01	0.28	327	4
22	0.11	0.28	328	4
23	0.21	0.27	330	3
24	0.32	0.27	331	3
25	0.42	0.27	332	3
26	0.52	0.27	334	3
27	0.62	0.27	335	3
28	0.73	0.27	336	3
29	0.84	0.27	338	3
30	0.96	0.28	339	4
31	1.08	0.28	341	4
32	1.21	0.29	342	4
33	1.36	0.30	344	4
34	1.51	0.32	346	4
35	1.69	0.34	348	4
36	1.90	0.37	351	5
37	2.16	0.42	354	5
38	2.48	0.49	358	6

Grade 3 Pretest Form M

Score	EAP	SD	Score	SE
0	-2.48	0.52	296	7
1	-2.40	0.52	297	7
2	-2.31	0.52	298	7
3	-2.21	0.53	299	7
4	-2.10	0.53	300	7
5	-1.98	0.53	302	7
6	-1.85	0.52	304	7
7	-1.71	0.51	305	6
8	-1.57	0.50	307	6
9	-1.42	0.48	309	6
10	-1.28	0.46	311	6
11	-1.14	0.43	313	5
12	-1.00	0.41	314	5
13	-0.87	0.39	316	5
14	-0.75	0.37	318	5
15	-0.62	0.35	319	4
16	-0.51	0.33	321	4
17	-0.40	0.32	322	4
18	-0.29	0.31	323	4
19	-0.18	0.30	325	4
20	-0.08	0.29	326	4
21	0.02	0.28	327	4
22	0.12	0.27	328	3
23	0.21	0.27	330	3
24	0.31	0.26	331	3
25	0.41	0.26	332	3
26	0.51	0.26	333	3
27	0.61	0.26	335	3
28	0.71	0.26	336	3
29	0.82	0.26	337	3
30	0.93	0.27	339	3
31	1.04	0.27	340	3
32	1.17	0.28	342	4
33	1.31	0.30	344	4
34	1.46	0.32	346	4
35	1.64	0.34	348	4
36	1.85	0.38	350	5
37	2.10	0.43	354	5
38	2.42	0.50	358	6

Grade 3 Pretest Form N

Score	EAP	SD	Score	SE
0	-2.67	0.52	293	7
1	-2.58	0.52	294	7
2	-2.47	0.53	296	7
3	-2.36	0.53	297	7
4	-2.24	0.53	299	7
5	-2.11	0.52	300	7
6	-1.97	0.52	302	7
7	-1.83	0.51	304	6
8	-1.69	0.49	306	6
9	-1.54	0.47	307	6
10	-1.40	0.45	309	6
11	-1.26	0.43	311	6
12	-1.12	0.42	313	5
13	-0.99	0.40	314	5
14	-0.86	0.38	316	5
15	-0.74	0.37	318	5
16	-0.62	0.35	319	4
17	-0.50	0.34	321	4
18	-0.39	0.33	322	4
19	-0.28	0.32	323	4
20	-0.17	0.31	325	4
21	-0.07	0.30	326	4
22	0.04	0.29	327	4
23	0.14	0.29	329	4
24	0.25	0.28	330	4
25	0.35	0.28	331	4
26	0.45	0.28	333	4
27	0.56	0.28	334	4
28	0.67	0.28	335	4
29	0.78	0.28	337	4
30	0.90	0.29	338	4
31	1.02	0.29	340	4
32	1.16	0.30	342	4
33	1.30	0.32	343	4
34	1.46	0.33	346	4
35	1.64	0.36	348	5
36	1.86	0.39	351	5
37	2.12	0.44	354	6
38	2.46	0.50	358	6

Grade 3 Pretest Form O

Score	EAP	SD	Score	SE
0	-2.62	0.52	294	7
1	-2.53	0.53	295	7
2	-2.45	0.53	296	7
3	-2.35	0.54	297	7
4	-2.24	0.54	299	7
5	-2.12	0.54	300	7
6	-2.00	0.54	302	7
7	-1.86	0.54	303	7
8	-1.72	0.53	305	7
9	-1.58	0.51	307	7
10	-1.43	0.50	309	6
11	-1.29	0.48	311	6
12	-1.15	0.46	312	6
13	-1.01	0.44	314	6
14	-0.87	0.42	316	5
15	-0.74	0.40	318	5
16	-0.62	0.39	319	5
17	-0.49	0.37	321	5
18	-0.38	0.35	322	4
19	-0.26	0.34	324	4
20	-0.15	0.32	325	4
21	-0.04	0.31	327	4
22	0.07	0.30	328	4
23	0.18	0.29	329	4
24	0.28	0.28	331	4
25	0.39	0.28	332	4
26	0.49	0.28	333	3
27	0.60	0.27	335	3
28	0.70	0.28	336	3
29	0.82	0.28	337	4
30	0.93	0.29	339	4
31	1.06	0.29	340	4
32	1.19	0.31	342	4
33	1.34	0.32	344	4
34	1.50	0.34	346	4
35	1.68	0.37	348	5
36	1.90	0.41	351	5
37	2.16	0.46	354	6
38	2.48	0.52	358	7

Grade 3 Pretest Form P

Score	EAP	SD	Score	SE
0	-2.56	0.52	295	7
1	-2.47	0.52	296	7
2	-2.38	0.52	297	7
3	-2.27	0.53	298	7
4	-2.16	0.53	300	7
5	-2.03	0.53	301	7
6	-1.90	0.52	303	7
7	-1.76	0.51	305	6
8	-1.61	0.49	307	6
9	-1.47	0.48	308	6
10	-1.32	0.46	310	6
11	-1.18	0.43	312	6
12	-1.04	0.41	314	5
13	-0.91	0.39	315	5
14	-0.79	0.38	317	5
15	-0.67	0.36	319	5
16	-0.55	0.34	320	4
17	-0.44	0.33	321	4
18	-0.33	0.32	323	4
19	-0.22	0.31	324	4
20	-0.11	0.30	326	4
21	-0.01	0.29	327	4
22	0.09	0.28	328	4
23	0.19	0.28	329	4
24	0.29	0.28	331	3
25	0.39	0.27	332	3
26	0.49	0.27	333	3
27	0.59	0.27	335	3
28	0.70	0.27	336	3
29	0.81	0.28	337	4
30	0.92	0.28	339	4
31	1.04	0.29	340	4
32	1.17	0.30	342	4
33	1.31	0.31	344	4
34	1.47	0.33	346	4
35	1.65	0.36	348	5
36	1.86	0.39	351	5
37	2.11	0.44	354	6
38	2.43	0.51	358	6

Grade 3 Form K

Score	EAP	SD	Score	SE
0	-2.53	0.52	312	6
1	-2.46	0.53	312	6
2	-2.39	0.53	313	6
3	-2.32	0.54	314	6
4	-2.24	0.54	315	6
5	-2.15	0.54	316	6
6	-2.06	0.54	317	6
7	-1.97	0.54	318	6
8	-1.86	0.54	319	6
9	-1.75	0.53	320	6
10	-1.64	0.52	321	6
11	-1.52	0.51	323	6
12	-1.40	0.50	324	5
13	-1.28	0.48	325	5
14	-1.15	0.46	327	5
15	-1.03	0.44	328	5
16	-0.92	0.42	329	5
17	-0.80	0.40	331	4
18	-0.69	0.38	332	4
19	-0.59	0.36	333	4
20	-0.49	0.34	334	4
21	-0.39	0.32	335	4
22	-0.29	0.31	336	3
23	-0.20	0.29	337	3
24	-0.12	0.28	338	3
25	-0.03	0.27	339	3
26	0.05	0.26	340	3
27	0.13	0.25	341	3
28	0.21	0.25	342	3
29	0.28	0.24	343	3
30	0.36	0.24	343	3
31	0.44	0.23	344	3
32	0.51	0.23	345	3
33	0.59	0.23	346	3
34	0.66	0.23	347	3
35	0.74	0.23	348	3
36	0.82	0.23	348	3
37	0.90	0.23	349	3
38	0.99	0.24	350	3
39	1.07	0.24	351	3
40	1.16	0.24	352	3
41	1.26	0.25	353	3
42	1.36	0.25	354	3
43	1.47	0.26	356	3
44	1.58	0.27	357	3
45	1.71	0.28	358	3
46	1.85	0.30	360	3
47	2.01	0.32	362	4
48	2.20	0.36	364	4
49	2.44	0.41	366	4
50	2.73	0.47	369	5

Grade 3 Form L

Score	EAP	SD	Score	SE
0	-2.55	0.52	312	6
1	-2.49	0.52	312	6
2	-2.42	0.53	313	6
3	-2.35	0.53	314	6
4	-2.27	0.53	315	6
5	-2.19	0.53	315	6
6	-2.10	0.54	316	6
7	-2.00	0.54	317	6
8	-1.90	0.53	319	6
9	-1.79	0.53	320	6
10	-1.68	0.52	321	6
11	-1.56	0.51	322	6
12	-1.44	0.50	324	5
13	-1.32	0.48	325	5
14	-1.20	0.46	326	5
15	-1.07	0.44	328	5
16	-0.96	0.42	329	5
17	-0.84	0.40	330	4
18	-0.73	0.38	331	4
19	-0.62	0.36	333	4
20	-0.52	0.34	334	4
21	-0.42	0.33	335	4
22	-0.32	0.31	336	3
23	-0.23	0.30	337	3
24	-0.14	0.29	338	3
25	-0.05	0.28	339	3
26	0.04	0.27	340	3
27	0.12	0.27	341	3
28	0.20	0.26	342	3
29	0.28	0.25	343	3
30	0.36	0.25	343	3
31	0.44	0.25	344	3
32	0.52	0.24	345	3
33	0.60	0.24	346	3
34	0.68	0.24	347	3
35	0.76	0.24	348	3
36	0.84	0.24	349	3
37	0.92	0.24	350	3
38	1.01	0.24	350	3
39	1.09	0.24	351	3
40	1.18	0.24	352	3
41	1.28	0.24	353	3
42	1.38	0.25	355	3
43	1.48	0.26	356	3
44	1.59	0.27	357	3
45	1.72	0.28	358	3
46	1.85	0.30	360	3
47	2.01	0.33	362	4
48	2.20	0.36	364	4
49	2.43	0.41	366	4
50	2.72	0.47	369	5

Grade 3 Form M

Score	EAP	SD	Score	SE
0	-2.54	0.51	312	6
1	-2.47	0.52	312	6
2	-2.40	0.52	313	6
3	-2.33	0.53	314	6
4	-2.25	0.53	315	6
5	-2.16	0.53	316	6
6	-2.06	0.53	317	6
7	-1.96	0.53	318	6
8	-1.85	0.53	319	6
9	-1.74	0.52	320	6
10	-1.62	0.51	322	6
11	-1.50	0.50	323	5
12	-1.37	0.48	324	5
13	-1.25	0.46	326	5
14	-1.13	0.44	327	5
15	-1.01	0.42	328	5
16	-0.89	0.40	330	4
17	-0.78	0.37	331	4
18	-0.68	0.36	332	4
19	-0.57	0.34	333	4
20	-0.48	0.32	334	4
21	-0.38	0.31	335	3
22	-0.29	0.29	336	3
23	-0.20	0.28	337	3
24	-0.12	0.27	338	3
25	-0.04	0.26	339	3
26	0.04	0.25	340	3
27	0.12	0.25	341	3
28	0.20	0.24	342	3
29	0.28	0.24	342	3
30	0.35	0.23	343	3
31	0.43	0.23	344	3
32	0.50	0.23	345	3
33	0.58	0.23	346	3
34	0.66	0.23	347	2
35	0.74	0.23	348	2
36	0.82	0.23	348	3
37	0.90	0.23	349	3
38	0.98	0.23	350	3
39	1.07	0.23	351	3
40	1.16	0.24	352	3
41	1.25	0.24	353	3
42	1.35	0.25	354	3
43	1.46	0.25	355	3
44	1.57	0.26	357	3
45	1.70	0.27	358	3
46	1.84	0.29	360	3
47	2.00	0.31	361	3
48	2.19	0.34	363	4
49	2.42	0.39	366	4
50	2.72	0.46	369	5

Grade 3 Form N

Score	EAP	SD	Score	SE
0	-2.63	0.52	311	6
1	-2.57	0.52	311	6
2	-2.50	0.53	312	6
3	-2.42	0.53	313	6
4	-2.34	0.54	314	6
5	-2.26	0.54	315	6
6	-2.16	0.54	316	6
7	-2.06	0.54	317	6
8	-1.96	0.54	318	6
9	-1.85	0.53	319	6
10	-1.73	0.53	320	6
11	-1.61	0.52	322	6
12	-1.49	0.51	323	6
13	-1.36	0.49	324	5
14	-1.24	0.47	326	5
15	-1.12	0.45	327	5
16	-1.00	0.44	329	5
17	-0.88	0.42	330	5
18	-0.76	0.40	331	4
19	-0.65	0.38	332	4
20	-0.54	0.36	333	4
21	-0.44	0.35	335	4
22	-0.34	0.33	336	4
23	-0.24	0.32	337	3
24	-0.15	0.30	338	3
25	-0.05	0.29	339	3
26	0.03	0.28	340	3
27	0.12	0.28	341	3
28	0.21	0.27	342	3
29	0.29	0.26	343	3
30	0.38	0.26	344	3
31	0.46	0.25	344	3
32	0.54	0.25	345	3
33	0.62	0.25	346	3
34	0.71	0.25	347	3
35	0.79	0.24	348	3
36	0.87	0.24	349	3
37	0.96	0.24	350	3
38	1.04	0.24	351	3
39	1.13	0.24	352	3
40	1.22	0.24	353	3
41	1.32	0.25	354	3
42	1.42	0.25	355	3
43	1.52	0.25	356	3
44	1.64	0.26	357	3
45	1.76	0.27	359	3
46	1.89	0.29	360	3
47	2.05	0.31	362	3
48	2.23	0.34	364	4
49	2.46	0.39	366	4
50	2.75	0.45	370	5

Grade 3 Form O

Score	EAP	SD	Score	SE
0	-2.58	0.52	311	6
1	-2.51	0.52	312	6
2	-2.45	0.53	313	6
3	-2.37	0.53	313	6
4	-2.29	0.54	314	6
5	-2.21	0.54	315	6
6	-2.11	0.54	316	6
7	-2.02	0.54	317	6
8	-1.91	0.54	318	6
9	-1.80	0.53	320	6
10	-1.68	0.53	321	6
11	-1.56	0.52	322	6
12	-1.44	0.50	324	6
13	-1.32	0.49	325	5
14	-1.19	0.47	326	5
15	-1.07	0.45	328	5
16	-0.95	0.43	329	5
17	-0.84	0.41	330	4
18	-0.72	0.39	332	4
19	-0.61	0.37	333	4
20	-0.51	0.35	334	4
21	-0.41	0.34	335	4
22	-0.31	0.32	336	4
23	-0.22	0.31	337	3
24	-0.12	0.30	338	3
25	-0.03	0.29	339	3
26	0.05	0.28	340	3
27	0.14	0.27	341	3
28	0.22	0.26	342	3
29	0.30	0.26	343	3
30	0.38	0.25	344	3
31	0.46	0.25	344	3
32	0.54	0.24	345	3
33	0.62	0.24	346	3
34	0.70	0.24	347	3
35	0.78	0.24	348	3
36	0.86	0.24	349	3
37	0.94	0.24	350	3
38	1.02	0.23	351	3
39	1.11	0.24	352	3
40	1.20	0.24	353	3
41	1.29	0.24	354	3
42	1.38	0.24	355	3
43	1.48	0.25	356	3
44	1.59	0.25	357	3
45	1.71	0.27	358	3
46	1.84	0.28	360	3
47	2.00	0.30	361	3
48	2.18	0.34	363	4
49	2.40	0.39	366	4
50	2.70	0.45	369	5

Grade 3 Form P

Score	EAP	SD	Score	SE
0	-2.60	0.51	311	6
1	-2.53	0.52	312	6
2	-2.46	0.52	312	6
3	-2.39	0.52	313	6
4	-2.31	0.53	314	6
5	-2.22	0.53	315	6
6	-2.12	0.53	316	6
7	-2.02	0.53	317	6
8	-1.91	0.53	318	6
9	-1.80	0.52	320	6
10	-1.68	0.51	321	6
11	-1.56	0.50	322	6
12	-1.43	0.49	324	5
13	-1.31	0.47	325	5
14	-1.18	0.45	326	5
15	-1.06	0.43	328	5
16	-0.94	0.41	329	5
17	-0.83	0.39	330	4
18	-0.72	0.37	332	4
19	-0.61	0.35	333	4
20	-0.51	0.34	334	4
21	-0.41	0.32	335	4
22	-0.32	0.31	336	3
23	-0.22	0.30	337	3
24	-0.14	0.29	338	3
25	-0.05	0.28	339	3
26	0.04	0.27	340	3
27	0.12	0.26	341	3
28	0.20	0.25	342	3
29	0.28	0.25	343	3
30	0.36	0.25	343	3
31	0.44	0.24	344	3
32	0.52	0.24	345	3
33	0.59	0.24	346	3
34	0.67	0.24	347	3
35	0.75	0.24	348	3
36	0.84	0.24	349	3
37	0.92	0.24	350	3
38	1.00	0.24	350	3
39	1.09	0.24	351	3
40	1.18	0.25	352	3
41	1.28	0.25	353	3
42	1.38	0.26	355	3
43	1.49	0.26	356	3
44	1.60	0.27	357	3
45	1.73	0.28	358	3
46	1.87	0.30	360	3
47	2.03	0.32	362	4
48	2.22	0.35	364	4
49	2.45	0.40	366	4
50	2.75	0.46	370	5

Grade 4 Form K

Score	EAP	SD	Score	SE
0	-2.49	0.53	320	5
1	-2.43	0.53	320	5
2	-2.36	0.54	321	5
3	-2.29	0.54	322	6
4	-2.22	0.54	323	6
5	-2.14	0.55	323	6
6	-2.06	0.55	324	6
7	-1.97	0.55	325	6
8	-1.87	0.55	326	6
9	-1.77	0.55	327	6
10	-1.66	0.54	328	6
11	-1.55	0.53	329	5
12	-1.43	0.52	331	5
13	-1.31	0.51	332	5
14	-1.18	0.49	333	5
15	-1.06	0.47	334	5
16	-0.94	0.45	336	5
17	-0.82	0.43	337	4
18	-0.70	0.41	338	4
19	-0.59	0.39	339	4
20	-0.48	0.37	340	4
21	-0.38	0.35	341	4
22	-0.28	0.33	342	3
23	-0.18	0.32	343	3
24	-0.09	0.30	344	3
25	0.00	0.29	345	3
26	0.09	0.28	346	3
27	0.17	0.27	347	3
28	0.26	0.26	348	3
29	0.34	0.26	349	3
30	0.41	0.25	349	3
31	0.49	0.24	350	3
32	0.57	0.24	351	2
33	0.65	0.24	352	2
34	0.72	0.23	353	2
35	0.80	0.23	353	2
36	0.88	0.23	354	2
37	0.95	0.23	355	2
38	1.03	0.23	356	2
39	1.12	0.23	357	2
40	1.20	0.23	358	2
41	1.29	0.24	358	2
42	1.38	0.24	359	2
43	1.48	0.25	360	3
44	1.59	0.26	362	3
45	1.70	0.27	363	3
46	1.84	0.29	364	3
47	1.99	0.31	366	3
48	2.17	0.35	368	4
49	2.40	0.39	370	4
50	2.70	0.46	373	5

Grade 4 Form L

Score	EAP	SD	Score	SE
0	-2.56	0.53	319	5
1	-2.48	0.54	320	5
2	-2.41	0.54	321	6
3	-2.33	0.54	321	6
4	-2.24	0.55	322	6
5	-2.15	0.55	323	6
6	-2.05	0.55	324	6
7	-1.95	0.55	325	6
8	-1.85	0.54	326	6
9	-1.74	0.54	327	5
10	-1.62	0.53	329	5
11	-1.50	0.52	330	5
12	-1.38	0.50	331	5
13	-1.26	0.48	332	5
14	-1.13	0.47	334	5
15	-1.01	0.44	335	5
16	-0.89	0.42	336	4
17	-0.78	0.40	337	4
18	-0.66	0.38	338	4
19	-0.56	0.36	340	4
20	-0.45	0.34	341	3
21	-0.36	0.32	342	3
22	-0.26	0.31	343	3
23	-0.17	0.29	344	3
24	-0.08	0.28	344	3
25	0.00	0.27	345	3
26	0.08	0.26	346	3
27	0.16	0.25	347	3
28	0.24	0.25	348	3
29	0.31	0.24	348	2
30	0.39	0.23	349	2
31	0.46	0.23	350	2
32	0.54	0.23	351	2
33	0.61	0.23	352	2
34	0.69	0.22	352	2
35	0.76	0.22	353	2
36	0.84	0.22	354	2
37	0.91	0.23	355	2
38	0.99	0.23	355	2
39	1.08	0.23	356	2
40	1.16	0.23	357	2
41	1.25	0.24	358	2
42	1.35	0.24	359	3
43	1.45	0.25	360	3
44	1.56	0.26	361	3
45	1.69	0.28	363	3
46	1.82	0.30	364	3
47	1.98	0.32	366	3
48	2.17	0.36	367	4
49	2.40	0.40	370	4
50	2.69	0.47	373	5

Grade 4 Form M

Score	EAP	SD	Score	SE
0	-2.40	0.52	321	5
1	-2.34	0.53	321	5
2	-2.28	0.53	322	5
3	-2.21	0.53	323	5
4	-2.14	0.54	323	5
5	-2.06	0.54	324	6
6	-1.98	0.54	325	6
7	-1.89	0.54	326	6
8	-1.79	0.54	327	5
9	-1.69	0.53	328	5
10	-1.58	0.52	329	5
11	-1.46	0.51	330	5
12	-1.34	0.50	332	5
13	-1.22	0.48	333	5
14	-1.10	0.46	334	5
15	-0.98	0.44	335	4
16	-0.86	0.41	336	4
17	-0.75	0.39	338	4
18	-0.64	0.37	339	4
19	-0.54	0.35	340	4
20	-0.44	0.33	341	3
21	-0.34	0.31	342	3
22	-0.25	0.30	343	3
23	-0.17	0.29	344	3
24	-0.08	0.28	344	3
25	-0.00	0.27	345	3
26	0.08	0.26	346	3
27	0.15	0.25	347	3
28	0.23	0.25	348	3
29	0.30	0.24	348	2
30	0.38	0.24	349	2
31	0.45	0.23	350	2
32	0.52	0.23	351	2
33	0.60	0.23	351	2
34	0.67	0.23	352	2
35	0.75	0.23	353	2
36	0.82	0.23	354	2
37	0.90	0.23	354	2
38	0.98	0.23	355	2
39	1.06	0.23	356	2
40	1.14	0.24	357	2
41	1.23	0.24	358	2
42	1.33	0.25	359	3
43	1.43	0.26	360	3
44	1.54	0.27	361	3
45	1.66	0.29	362	3
46	1.80	0.30	364	3
47	1.96	0.33	365	3
48	2.15	0.36	367	4
49	2.38	0.41	370	4
50	2.68	0.48	373	5

Grade 4 Form N

Score	EAP	SD	Score	SE
0	-2.57	0.52	319	5
1	-2.51	0.52	320	5
2	-2.43	0.53	320	5
3	-2.36	0.53	321	5
4	-2.27	0.54	322	5
5	-2.18	0.54	323	6
6	-2.08	0.54	324	6
7	-1.98	0.54	325	6
8	-1.86	0.54	326	5
9	-1.75	0.53	327	5
10	-1.63	0.52	329	5
11	-1.50	0.51	330	5
12	-1.37	0.49	331	5
13	-1.25	0.47	332	5
14	-1.12	0.46	334	5
15	-1.00	0.43	335	4
16	-0.88	0.41	336	4
17	-0.77	0.39	337	4
18	-0.66	0.37	339	4
19	-0.55	0.36	340	4
20	-0.45	0.34	341	3
21	-0.35	0.32	342	3
22	-0.25	0.31	343	3
23	-0.16	0.30	344	3
24	-0.08	0.28	344	3
25	0.01	0.27	345	3
26	0.09	0.26	346	3
27	0.17	0.25	347	3
28	0.25	0.25	348	3
29	0.32	0.24	349	2
30	0.40	0.23	349	2
31	0.47	0.23	350	2
32	0.54	0.22	351	2
33	0.61	0.22	352	2
34	0.68	0.22	352	2
35	0.76	0.22	353	2
36	0.83	0.22	354	2
37	0.90	0.22	355	2
38	0.98	0.22	355	2
39	1.06	0.22	356	2
40	1.14	0.23	357	2
41	1.23	0.24	358	2
42	1.32	0.24	359	2
43	1.42	0.25	360	3
44	1.53	0.27	361	3
45	1.65	0.29	362	3
46	1.79	0.31	364	3
47	1.95	0.34	365	3
48	2.14	0.37	367	4
49	2.37	0.42	370	4
50	2.65	0.48	372	5

Grade 4 Form O

Score	EAP	SD	Score	SE
0	-2.53	0.54	319	5
1	-2.47	0.54	320	6
2	-2.39	0.54	321	6
3	-2.32	0.55	322	6
4	-2.24	0.55	322	6
5	-2.15	0.55	323	6
6	-2.06	0.55	324	6
7	-1.97	0.55	325	6
8	-1.87	0.55	326	6
9	-1.76	0.55	327	6
10	-1.65	0.54	328	6
11	-1.53	0.53	330	5
12	-1.41	0.52	331	5
13	-1.29	0.51	332	5
14	-1.17	0.49	333	5
15	-1.05	0.47	335	5
16	-0.92	0.45	336	5
17	-0.80	0.43	337	4
18	-0.69	0.40	338	4
19	-0.58	0.38	339	4
20	-0.47	0.36	340	4
21	-0.37	0.34	341	3
22	-0.27	0.32	343	3
23	-0.17	0.31	343	3
24	-0.08	0.29	344	3
25	0.00	0.28	345	3
26	0.09	0.27	346	3
27	0.17	0.26	347	3
28	0.25	0.25	348	3
29	0.32	0.24	349	2
30	0.40	0.24	349	2
31	0.48	0.23	350	2
32	0.55	0.23	351	2
33	0.62	0.22	352	2
34	0.70	0.22	352	2
35	0.77	0.22	353	2
36	0.85	0.22	354	2
37	0.92	0.22	355	2
38	1.00	0.22	356	2
39	1.08	0.22	356	2
40	1.17	0.23	357	2
41	1.26	0.23	358	2
42	1.35	0.24	359	2
43	1.45	0.25	360	3
44	1.56	0.26	361	3
45	1.68	0.27	362	3
46	1.82	0.29	364	3
47	1.98	0.32	365	3
48	2.16	0.35	367	4
49	2.39	0.40	370	4
50	2.68	0.47	373	5

Grade 4 Form P

Score	EAP	SD	Score	SE
0	-2.56	0.52	319	5
1	-2.49	0.52	320	5
2	-2.42	0.53	320	5
3	-2.35	0.53	321	5
4	-2.26	0.54	322	6
5	-2.17	0.54	323	6
6	-2.08	0.54	324	6
7	-1.98	0.54	325	6
8	-1.87	0.54	326	6
9	-1.76	0.53	327	5
10	-1.64	0.53	328	5
11	-1.52	0.51	330	5
12	-1.39	0.50	331	5
13	-1.27	0.48	332	5
14	-1.15	0.46	334	5
15	-1.02	0.44	335	5
16	-0.91	0.42	336	4
17	-0.79	0.40	337	4
18	-0.68	0.38	338	4
19	-0.57	0.36	339	4
20	-0.47	0.35	340	4
21	-0.37	0.33	341	3
22	-0.27	0.31	342	3
23	-0.18	0.30	343	3
24	-0.09	0.29	344	3
25	-0.01	0.28	345	3
26	0.08	0.27	346	3
27	0.16	0.26	347	3
28	0.24	0.25	348	3
29	0.32	0.25	348	3
30	0.39	0.24	349	2
31	0.47	0.24	350	2
32	0.55	0.24	351	2
33	0.62	0.23	352	2
34	0.70	0.23	352	2
35	0.78	0.23	353	2
36	0.86	0.23	354	2
37	0.94	0.23	355	2
38	1.02	0.24	356	2
39	1.11	0.24	357	2
40	1.20	0.25	358	3
41	1.29	0.25	358	3
42	1.39	0.26	360	3
43	1.50	0.27	361	3
44	1.62	0.28	362	3
45	1.75	0.30	363	3
46	1.89	0.32	365	3
47	2.06	0.35	366	4
48	2.25	0.38	368	4
49	2.49	0.42	371	4
50	2.78	0.48	374	5

Grade 5 Form K

Score	EAP	SD	Score	SE
0	-2.44	0.52	326	5
1	-2.38	0.53	326	5
2	-2.32	0.53	327	5
3	-2.26	0.53	327	5
4	-2.19	0.54	328	5
5	-2.11	0.54	329	5
6	-2.03	0.54	330	5
7	-1.95	0.54	331	5
8	-1.85	0.54	331	5
9	-1.75	0.54	332	5
10	-1.65	0.54	334	5
11	-1.53	0.53	335	5
12	-1.42	0.52	336	5
13	-1.30	0.50	337	5
14	-1.17	0.49	338	5
15	-1.05	0.47	339	5
16	-0.93	0.45	341	4
17	-0.81	0.42	342	4
18	-0.70	0.40	343	4
19	-0.59	0.38	344	4
20	-0.48	0.36	345	4
21	-0.38	0.35	346	3
22	-0.28	0.33	347	3
23	-0.18	0.32	348	3
24	-0.09	0.31	349	3
25	-0.00	0.29	350	3
26	0.08	0.28	351	3
27	0.17	0.28	352	3
28	0.25	0.27	352	3
29	0.33	0.26	353	3
30	0.41	0.26	354	3
31	0.49	0.25	355	3
32	0.57	0.25	356	2
33	0.65	0.24	356	2
34	0.72	0.24	357	2
35	0.80	0.24	358	2
36	0.88	0.24	359	2
37	0.96	0.23	360	2
38	1.05	0.24	360	2
39	1.13	0.24	361	2
40	1.22	0.24	362	2
41	1.31	0.24	363	2
42	1.41	0.25	364	2
43	1.51	0.26	365	3
44	1.62	0.27	366	3
45	1.75	0.28	367	3
46	1.89	0.31	369	3
47	2.05	0.33	370	3
48	2.24	0.37	372	4
49	2.47	0.42	375	4
50	2.75	0.48	377	5

Grade 5 Form L

Score	EAP	SD	Score	SE
0	-2.29	0.53	327	5
1	-2.24	0.53	328	5
2	-2.19	0.54	328	5
3	-2.13	0.54	329	5
4	-2.07	0.55	329	5
5	-2.00	0.55	330	5
6	-1.93	0.55	331	6
7	-1.85	0.55	332	6
8	-1.76	0.55	332	6
9	-1.68	0.55	333	6
10	-1.58	0.55	334	5
11	-1.48	0.54	335	5
12	-1.37	0.53	336	5
13	-1.26	0.52	337	5
14	-1.14	0.51	339	5
15	-1.02	0.49	340	5
16	-0.91	0.47	341	5
17	-0.79	0.44	342	4
18	-0.67	0.42	343	4
19	-0.56	0.39	344	4
20	-0.45	0.37	345	4
21	-0.35	0.34	346	3
22	-0.25	0.32	347	3
23	-0.16	0.31	348	3
24	-0.07	0.29	349	3
25	0.01	0.28	350	3
26	0.09	0.27	351	3
27	0.17	0.26	352	3
28	0.25	0.25	352	3
29	0.32	0.24	353	2
30	0.40	0.24	354	2
31	0.47	0.24	355	2
32	0.54	0.23	355	2
33	0.62	0.23	356	2
34	0.69	0.23	357	2
35	0.76	0.23	358	2
36	0.84	0.23	358	2
37	0.92	0.23	359	2
38	1.00	0.23	360	2
39	1.08	0.23	361	2
40	1.17	0.24	362	2
41	1.26	0.25	363	2
42	1.36	0.26	364	3
43	1.47	0.27	365	3
44	1.58	0.28	366	3
45	1.72	0.30	367	3
46	1.87	0.33	369	3
47	2.04	0.37	370	4
48	2.25	0.41	372	4
49	2.48	0.46	375	5
50	2.74	0.51	377	5

Grade 5 Form M

Score	EAP	SD	Score	SE
0	-2.44	0.52	326	5
1	-2.39	0.52	326	5
2	-2.33	0.53	327	5
3	-2.27	0.53	327	5
4	-2.20	0.54	328	5
5	-2.12	0.54	329	5
6	-2.05	0.54	330	5
7	-1.96	0.55	330	5
8	-1.87	0.55	331	5
9	-1.77	0.55	332	5
10	-1.66	0.54	333	5
11	-1.55	0.54	335	5
12	-1.43	0.53	336	5
13	-1.31	0.51	337	5
14	-1.19	0.50	338	5
15	-1.07	0.48	339	5
16	-0.95	0.46	341	5
17	-0.83	0.43	342	4
18	-0.71	0.41	343	4
19	-0.60	0.39	344	4
20	-0.49	0.37	345	4
21	-0.39	0.35	346	4
22	-0.29	0.34	347	3
23	-0.19	0.32	348	3
24	-0.10	0.31	349	3
25	-0.01	0.30	350	3
26	0.08	0.29	351	3
27	0.16	0.28	352	3
28	0.25	0.27	352	3
29	0.33	0.26	353	3
30	0.41	0.26	354	3
31	0.49	0.25	355	3
32	0.56	0.25	356	2
33	0.64	0.25	356	2
34	0.72	0.24	357	2
35	0.80	0.24	358	2
36	0.88	0.24	359	2
37	0.96	0.24	360	2
38	1.05	0.24	360	2
39	1.13	0.25	361	2
40	1.23	0.25	362	2
41	1.32	0.25	363	3
42	1.42	0.26	364	3
43	1.53	0.27	365	3
44	1.65	0.29	366	3
45	1.78	0.30	368	3
46	1.92	0.32	369	3
47	2.09	0.35	371	4
48	2.28	0.39	373	4
49	2.51	0.44	375	4
50	2.79	0.49	378	5

Grade 5 Form N

Score	EAP	SD	Score	SE
0	-2.25	0.53	327	5
1	-2.21	0.54	328	5
2	-2.15	0.54	328	5
3	-2.10	0.54	329	5
4	-2.04	0.55	330	5
5	-1.97	0.55	330	6
6	-1.90	0.55	331	6
7	-1.83	0.56	332	6
8	-1.75	0.56	333	6
9	-1.66	0.56	333	6
10	-1.57	0.56	334	6
11	-1.47	0.55	335	6
12	-1.36	0.54	336	5
13	-1.25	0.53	337	5
14	-1.14	0.52	339	5
15	-1.02	0.50	340	5
16	-0.90	0.48	341	5
17	-0.78	0.46	342	5
18	-0.67	0.43	343	4
19	-0.55	0.40	344	4
20	-0.44	0.38	346	4
21	-0.34	0.35	347	4
22	-0.24	0.33	348	3
23	-0.15	0.31	349	3
24	-0.06	0.29	349	3
25	0.03	0.28	350	3
26	0.11	0.27	351	3
27	0.19	0.25	352	3
28	0.26	0.25	353	2
29	0.34	0.24	353	2
30	0.41	0.23	354	2
31	0.48	0.23	355	2
32	0.55	0.23	356	2
33	0.62	0.22	356	2
34	0.69	0.22	357	2
35	0.76	0.22	358	2
36	0.84	0.22	358	2
37	0.91	0.22	359	2
38	0.99	0.22	360	2
39	1.07	0.23	361	2
40	1.15	0.23	362	2
41	1.24	0.24	362	2
42	1.33	0.24	363	2
43	1.43	0.26	364	3
44	1.55	0.27	365	3
45	1.67	0.29	367	3
46	1.82	0.31	368	3
47	1.98	0.35	370	3
48	2.18	0.39	372	4
49	2.41	0.44	374	4
50	2.68	0.50	377	5

Grade 5 Form O

Score	EAP	SD	Score	SE
0	-2.41	0.53	326	5
1	-2.35	0.53	327	5
2	-2.28	0.54	327	5
3	-2.22	0.54	328	5
4	-2.15	0.54	329	5
5	-2.07	0.55	329	5
6	-1.99	0.55	330	5
7	-1.90	0.55	331	5
8	-1.81	0.55	332	5
9	-1.71	0.55	333	5
10	-1.60	0.54	334	5
11	-1.49	0.53	335	5
12	-1.38	0.52	336	5
13	-1.26	0.51	337	5
14	-1.14	0.49	339	5
15	-1.02	0.47	340	5
16	-0.90	0.45	341	5
17	-0.78	0.43	342	4
18	-0.67	0.40	343	4
19	-0.56	0.38	344	4
20	-0.45	0.36	345	4
21	-0.35	0.34	346	3
22	-0.26	0.32	347	3
23	-0.17	0.30	348	3
24	-0.08	0.29	349	3
25	0.01	0.28	350	3
26	0.09	0.27	351	3
27	0.17	0.26	352	3
28	0.25	0.25	352	2
29	0.32	0.24	353	2
30	0.40	0.24	354	2
31	0.47	0.23	355	2
32	0.54	0.23	355	2
33	0.62	0.23	356	2
34	0.69	0.23	357	2
35	0.77	0.23	358	2
36	0.84	0.23	358	2
37	0.92	0.23	359	2
38	1.00	0.23	360	2
39	1.09	0.24	361	2
40	1.17	0.24	362	2
41	1.27	0.25	363	2
42	1.36	0.25	364	3
43	1.47	0.26	365	3
44	1.59	0.28	366	3
45	1.72	0.29	367	3
46	1.86	0.32	369	3
47	2.03	0.34	370	3
48	2.23	0.38	372	4
49	2.47	0.43	375	4
50	2.75	0.48	377	5

Grade 5 Form P

Score	EAP	SD	Score	SE
0	-2.33	0.52	327	5
1	-2.28	0.52	327	5
2	-2.22	0.53	328	5
3	-2.15	0.53	328	5
4	-2.09	0.53	329	5
5	-2.02	0.53	330	5
6	-1.94	0.54	331	5
7	-1.86	0.54	331	5
8	-1.77	0.54	332	5
9	-1.67	0.53	333	5
10	-1.57	0.53	334	5
11	-1.46	0.52	335	5
12	-1.35	0.51	337	5
13	-1.23	0.49	338	5
14	-1.11	0.47	339	5
15	-0.99	0.45	340	5
16	-0.87	0.43	341	4
17	-0.76	0.40	342	4
18	-0.65	0.38	344	4
19	-0.54	0.36	345	4
20	-0.44	0.33	346	3
21	-0.35	0.32	347	3
22	-0.26	0.30	347	3
23	-0.17	0.29	348	3
24	-0.08	0.28	349	3
25	-0.00	0.27	350	3
26	0.08	0.26	351	3
27	0.15	0.25	352	3
28	0.23	0.25	352	2
29	0.31	0.24	353	2
30	0.38	0.24	354	2
31	0.45	0.23	355	2
32	0.53	0.23	355	2
33	0.60	0.23	356	2
34	0.68	0.23	357	2
35	0.75	0.23	358	2
36	0.83	0.23	358	2
37	0.91	0.23	359	2
38	0.99	0.23	360	2
39	1.07	0.23	361	2
40	1.16	0.24	362	2
41	1.25	0.24	363	2
42	1.35	0.25	364	3
43	1.46	0.26	365	3
44	1.57	0.27	366	3
45	1.70	0.28	367	3
46	1.84	0.30	368	3
47	2.00	0.33	370	3
48	2.19	0.36	372	4
49	2.43	0.41	374	4
50	2.74	0.47	377	5

Grade 6 Form K

Score	EAP	SD	Score	SE
0	-2.20	0.53	329	6
1	-2.15	0.54	329	6
2	-2.09	0.54	330	6
3	-2.03	0.55	330	6
4	-1.97	0.55	331	6
5	-1.91	0.56	332	6
6	-1.84	0.56	332	6
7	-1.76	0.56	333	6
8	-1.68	0.56	334	6
9	-1.59	0.56	335	6
10	-1.49	0.56	336	6
11	-1.39	0.56	337	6
12	-1.29	0.55	338	6
13	-1.17	0.54	339	6
14	-1.05	0.53	340	6
15	-0.93	0.51	342	5
16	-0.81	0.49	343	5
17	-0.68	0.46	344	5
18	-0.56	0.44	346	5
19	-0.44	0.41	347	4
20	-0.33	0.38	348	4
21	-0.22	0.36	349	4
22	-0.12	0.34	350	3
23	-0.02	0.32	351	3
24	0.07	0.30	352	3
25	0.16	0.28	353	3
26	0.24	0.27	354	3
27	0.32	0.26	355	3
28	0.40	0.25	356	3
29	0.48	0.24	356	3
30	0.56	0.24	357	2
31	0.63	0.23	358	2
32	0.70	0.23	359	2
33	0.78	0.22	360	2
34	0.85	0.22	360	2
35	0.92	0.21	361	2
36	0.99	0.21	362	2
37	1.07	0.21	363	2
38	1.14	0.21	363	2
39	1.22	0.21	364	2
40	1.29	0.21	365	2
41	1.37	0.21	366	2
42	1.46	0.21	367	2
43	1.55	0.22	368	2
44	1.65	0.23	369	2
45	1.75	0.24	370	2
46	1.87	0.26	371	3
47	2.01	0.28	372	3
48	2.18	0.32	374	3
49	2.40	0.37	376	4
50	2.70	0.45	380	5

Grade 6 Form L

Score	EAP	SD	Score	SE
0	-2.19	0.52	329	5
1	-2.14	0.52	329	5
2	-2.08	0.53	330	5
3	-2.03	0.53	330	6
4	-1.97	0.53	331	6
5	-1.90	0.54	332	6
6	-1.82	0.54	332	6
7	-1.74	0.54	333	6
8	-1.66	0.54	334	6
9	-1.56	0.54	335	6
10	-1.46	0.54	336	6
11	-1.35	0.53	337	6
12	-1.23	0.52	339	5
13	-1.11	0.50	340	5
14	-0.99	0.48	341	5
15	-0.86	0.46	342	5
16	-0.74	0.43	344	4
17	-0.63	0.40	345	4
18	-0.51	0.38	346	4
19	-0.41	0.35	347	4
20	-0.30	0.33	348	3
21	-0.21	0.32	349	3
22	-0.12	0.30	350	3
23	-0.03	0.29	351	3
24	0.06	0.28	352	3
25	0.14	0.27	353	3
26	0.22	0.26	354	3
27	0.29	0.25	355	3
28	0.37	0.25	355	3
29	0.45	0.24	356	3
30	0.52	0.24	357	2
31	0.59	0.23	358	2
32	0.67	0.23	358	2
33	0.74	0.23	359	2
34	0.82	0.22	360	2
35	0.89	0.22	361	2
36	0.97	0.22	362	2
37	1.04	0.22	362	2
38	1.12	0.22	363	2
39	1.20	0.22	364	2
40	1.28	0.22	365	2
41	1.37	0.22	366	2
42	1.46	0.23	367	2
43	1.56	0.23	368	2
44	1.66	0.24	369	3
45	1.78	0.25	370	3
46	1.90	0.27	371	3
47	2.05	0.29	373	3
48	2.23	0.32	375	3
49	2.46	0.37	377	4
50	2.76	0.44	380	5

Grade 6 Form M

Score	EAP	SD	Score	SE
0	-2.21	0.53	328	6
1	-2.16	0.54	329	6
2	-2.10	0.54	330	6
3	-2.04	0.55	330	6
4	-1.98	0.55	331	6
5	-1.91	0.55	332	6
6	-1.83	0.56	332	6
7	-1.75	0.56	333	6
8	-1.66	0.56	334	6
9	-1.57	0.56	335	6
10	-1.47	0.55	336	6
11	-1.36	0.55	337	6
12	-1.25	0.54	338	6
13	-1.13	0.53	340	5
14	-1.01	0.51	341	5
15	-0.89	0.49	342	5
16	-0.77	0.46	343	5
17	-0.65	0.44	345	5
18	-0.53	0.41	346	4
19	-0.42	0.38	347	4
20	-0.31	0.36	348	4
21	-0.21	0.33	349	3
22	-0.11	0.31	350	3
23	-0.02	0.30	351	3
24	0.07	0.28	352	3
25	0.15	0.27	353	3
26	0.23	0.26	354	3
27	0.31	0.25	355	3
28	0.39	0.24	355	3
29	0.46	0.24	356	2
30	0.54	0.23	357	2
31	0.61	0.23	358	2
32	0.68	0.22	359	2
33	0.75	0.22	359	2
34	0.83	0.22	360	2
35	0.90	0.21	361	2
36	0.97	0.21	362	2
37	1.04	0.21	362	2
38	1.12	0.21	363	2
39	1.20	0.21	364	2
40	1.28	0.21	365	2
41	1.36	0.22	366	2
42	1.45	0.22	367	2
43	1.54	0.23	368	2
44	1.65	0.24	369	2
45	1.76	0.25	370	3
46	1.89	0.27	371	3
47	2.04	0.30	373	3
48	2.22	0.34	375	4
49	2.45	0.39	377	4
50	2.72	0.46	380	5

Grade 6 Form N

Score	EAP	SD	Score	SE
0	-2.26	0.52	328	5
1	-2.21	0.53	328	5
2	-2.16	0.53	329	6
3	-2.10	0.54	330	6
4	-2.03	0.54	330	6
5	-1.96	0.55	331	6
6	-1.89	0.55	332	6
7	-1.80	0.55	333	6
8	-1.71	0.55	334	6
9	-1.62	0.55	335	6
10	-1.51	0.55	336	6
11	-1.40	0.55	337	6
12	-1.29	0.54	338	6
13	-1.16	0.52	339	5
14	-1.04	0.51	341	5
15	-0.92	0.49	342	5
16	-0.79	0.46	343	5
17	-0.67	0.44	344	5
18	-0.55	0.41	346	4
19	-0.44	0.39	347	4
20	-0.33	0.36	348	4
21	-0.23	0.34	349	4
22	-0.13	0.32	350	3
23	-0.03	0.31	351	3
24	0.06	0.29	352	3
25	0.14	0.28	353	3
26	0.23	0.27	354	3
27	0.31	0.26	355	3
28	0.39	0.25	356	3
29	0.47	0.25	356	3
30	0.54	0.24	357	2
31	0.62	0.23	358	2
32	0.69	0.23	359	2
33	0.77	0.23	359	2
34	0.84	0.23	360	2
35	0.92	0.22	361	2
36	0.99	0.22	362	2
37	1.07	0.22	363	2
38	1.15	0.22	363	2
39	1.23	0.22	364	2
40	1.31	0.22	365	2
41	1.39	0.22	366	2
42	1.48	0.23	367	2
43	1.58	0.23	368	2
44	1.68	0.24	369	2
45	1.79	0.25	370	3
46	1.92	0.27	371	3
47	2.06	0.29	373	3
48	2.24	0.33	375	3
49	2.46	0.38	377	4
50	2.74	0.45	380	5

Grade 6 Form O

Score	EAP	SD	Score	SE
0	-2.38	0.52	327	5
1	-2.32	0.53	327	5
2	-2.26	0.53	328	6
3	-2.19	0.54	329	6
4	-2.11	0.54	329	6
5	-2.03	0.54	330	6
6	-1.94	0.55	331	6
7	-1.85	0.55	332	6
8	-1.75	0.55	333	6
9	-1.64	0.54	334	6
10	-1.53	0.54	336	6
11	-1.41	0.53	337	6
12	-1.29	0.52	338	5
13	-1.16	0.50	339	5
14	-1.03	0.48	341	5
15	-0.91	0.46	342	5
16	-0.78	0.44	343	5
17	-0.66	0.41	345	4
18	-0.55	0.39	346	4
19	-0.44	0.37	347	4
20	-0.33	0.35	348	4
21	-0.23	0.33	349	3
22	-0.13	0.32	350	3
23	-0.04	0.30	351	3
24	0.05	0.29	352	3
25	0.14	0.28	353	3
26	0.22	0.27	354	3
27	0.31	0.26	355	3
28	0.39	0.25	355	3
29	0.47	0.25	356	3
30	0.55	0.24	357	3
31	0.62	0.24	358	2
32	0.70	0.23	359	2
33	0.78	0.23	360	2
34	0.85	0.23	360	2
35	0.93	0.23	361	2
36	1.01	0.22	362	2
37	1.09	0.22	363	2
38	1.17	0.22	364	2
39	1.25	0.22	364	2
40	1.34	0.22	365	2
41	1.43	0.23	366	2
42	1.52	0.23	367	2
43	1.62	0.23	368	2
44	1.72	0.24	369	2
45	1.84	0.25	371	3
46	1.97	0.26	372	3
47	2.11	0.28	373	3
48	2.29	0.31	375	3
49	2.51	0.36	378	4
50	2.80	0.44	381	5

Grade 6 Form P

Score	EAP	SD	Score	SE
0	-2.31	0.53	327	6
1	-2.25	0.53	328	6
2	-2.19	0.54	329	6
3	-2.12	0.54	329	6
4	-2.05	0.55	330	6
5	-1.97	0.55	331	6
6	-1.89	0.55	332	6
7	-1.80	0.55	333	6
8	-1.70	0.55	334	6
9	-1.60	0.55	335	6
10	-1.49	0.54	336	6
11	-1.38	0.53	337	6
12	-1.26	0.52	338	5
13	-1.13	0.51	340	5
14	-1.01	0.49	341	5
15	-0.89	0.46	342	5
16	-0.76	0.44	344	5
17	-0.64	0.42	345	4
18	-0.53	0.39	346	4
19	-0.42	0.37	347	4
20	-0.31	0.35	348	4
21	-0.21	0.33	349	3
22	-0.12	0.31	350	3
23	-0.03	0.30	351	3
24	0.06	0.28	352	3
25	0.14	0.27	353	3
26	0.23	0.26	354	3
27	0.30	0.25	355	3
28	0.38	0.25	355	3
29	0.46	0.24	356	2
30	0.53	0.23	357	2
31	0.61	0.23	358	2
32	0.68	0.23	359	2
33	0.75	0.22	359	2
34	0.83	0.22	360	2
35	0.90	0.22	361	2
36	0.97	0.22	362	2
37	1.05	0.22	362	2
38	1.13	0.22	363	2
39	1.21	0.22	364	2
40	1.29	0.23	365	2
41	1.38	0.23	366	2
42	1.47	0.24	367	2
43	1.57	0.25	368	3
44	1.68	0.26	369	3
45	1.80	0.28	370	3
46	1.94	0.30	372	3
47	2.09	0.33	373	3
48	2.28	0.36	375	4
49	2.51	0.41	378	4
50	2.78	0.47	380	5

Grade 7 Form K

Score	EAP	SD	Score	SE
0	-2.10	0.52	332	5
1	-2.06	0.53	333	5
2	-2.01	0.53	333	5
3	-1.95	0.54	334	5
4	-1.90	0.54	334	6
5	-1.83	0.55	335	6
6	-1.76	0.55	336	6
7	-1.68	0.55	337	6
8	-1.60	0.55	337	6
9	-1.51	0.55	338	6
10	-1.41	0.55	339	6
11	-1.30	0.54	340	6
12	-1.18	0.53	342	5
13	-1.06	0.51	343	5
14	-0.94	0.49	344	5
15	-0.82	0.47	345	5
16	-0.70	0.44	347	4
17	-0.58	0.41	348	4
18	-0.46	0.38	349	4
19	-0.35	0.36	350	4
20	-0.25	0.33	351	3
21	-0.16	0.31	352	3
22	-0.06	0.30	353	3
23	0.02	0.28	354	3
24	0.11	0.27	355	3
25	0.19	0.26	356	3
26	0.27	0.25	356	3
27	0.34	0.25	357	3
28	0.42	0.24	358	2
29	0.49	0.24	359	2
30	0.57	0.23	359	2
31	0.64	0.23	360	2
32	0.71	0.23	361	2
33	0.79	0.22	362	2
34	0.86	0.22	362	2
35	0.93	0.22	363	2
36	1.01	0.22	364	2
37	1.08	0.22	365	2
38	1.16	0.22	365	2
39	1.24	0.22	366	2
40	1.33	0.23	367	2
41	1.42	0.23	368	2
42	1.51	0.24	369	2
43	1.61	0.24	370	2
44	1.71	0.25	371	3
45	1.83	0.27	372	3
46	1.96	0.29	374	3
47	2.11	0.31	375	3
48	2.30	0.34	377	3
49	2.52	0.39	379	4
50	2.81	0.46	382	5

Grade 7 Form L

Score	EAP	SD	Score	SE
0	-2.17	0.53	332	5
1	-2.12	0.53	332	5
2	-2.06	0.54	333	5
3	-2.00	0.54	333	5
4	-1.94	0.55	334	6
5	-1.86	0.55	335	6
6	-1.78	0.55	336	6
7	-1.69	0.55	336	6
8	-1.60	0.55	337	6
9	-1.50	0.55	338	6
10	-1.39	0.54	340	6
11	-1.27	0.53	341	5
12	-1.15	0.52	342	5
13	-1.03	0.50	343	5
14	-0.91	0.48	344	5
15	-0.78	0.45	346	5
16	-0.66	0.42	347	4
17	-0.55	0.40	348	4
18	-0.44	0.37	349	4
19	-0.33	0.35	350	4
20	-0.24	0.33	351	3
21	-0.14	0.31	352	3
22	-0.05	0.29	353	3
23	0.03	0.28	354	3
24	0.11	0.27	355	3
25	0.19	0.26	356	3
26	0.27	0.25	356	3
27	0.34	0.24	357	2
28	0.41	0.24	358	2
29	0.49	0.23	359	2
30	0.56	0.23	359	2
31	0.63	0.22	360	2
32	0.70	0.22	361	2
33	0.77	0.22	361	2
34	0.84	0.22	362	2
35	0.91	0.22	363	2
36	0.98	0.22	364	2
37	1.05	0.22	364	2
38	1.13	0.22	365	2
39	1.21	0.22	366	2
40	1.29	0.23	367	2
41	1.38	0.23	368	2
42	1.47	0.24	369	2
43	1.57	0.25	370	3
44	1.68	0.26	371	3
45	1.80	0.28	372	3
46	1.93	0.30	373	3
47	2.09	0.32	375	3
48	2.27	0.36	377	4
49	2.50	0.40	379	4
50	2.79	0.47	382	5

Grade 7 Form M

Score	EAP	SD	Score	SE
0	-2.09	0.53	332	5
1	-2.04	0.53	333	5
2	-1.99	0.54	333	5
3	-1.93	0.54	334	5
4	-1.86	0.54	335	6
5	-1.79	0.55	335	6
6	-1.71	0.55	336	6
7	-1.63	0.55	337	6
8	-1.54	0.55	338	6
9	-1.44	0.54	339	6
10	-1.33	0.54	340	5
11	-1.22	0.52	341	5
12	-1.10	0.51	342	5
13	-0.98	0.49	344	5
14	-0.86	0.46	345	5
15	-0.74	0.43	346	4
16	-0.62	0.40	347	4
17	-0.51	0.37	349	4
18	-0.40	0.34	350	3
19	-0.31	0.32	351	3
20	-0.21	0.30	351	3
21	-0.13	0.28	352	3
22	-0.04	0.27	353	3
23	0.03	0.26	354	3
24	0.11	0.25	355	3
25	0.18	0.24	356	2
26	0.25	0.23	356	2
27	0.32	0.23	357	2
28	0.39	0.22	358	2
29	0.46	0.22	358	2
30	0.53	0.22	359	2
31	0.60	0.21	360	2
32	0.66	0.21	360	2
33	0.73	0.21	361	2
34	0.80	0.21	362	2
35	0.87	0.21	363	2
36	0.94	0.21	363	2
37	1.02	0.21	364	2
38	1.09	0.22	365	2
39	1.17	0.22	366	2
40	1.25	0.22	366	2
41	1.34	0.23	367	2
42	1.43	0.24	368	2
43	1.52	0.24	369	2
44	1.63	0.25	370	3
45	1.75	0.27	371	3
46	1.88	0.29	373	3
47	2.03	0.31	374	3
48	2.22	0.35	376	4
49	2.45	0.39	378	4
50	2.75	0.46	382	5

Grade 7 Form N

Score	EAP	SD	Score	SE
0	-2.07	0.53	333	5
1	-2.02	0.53	333	5
2	-1.97	0.54	334	5
3	-1.91	0.54	334	6
4	-1.85	0.55	335	6
5	-1.79	0.55	335	6
6	-1.72	0.55	336	6
7	-1.64	0.55	337	6
8	-1.55	0.55	338	6
9	-1.46	0.55	339	6
10	-1.36	0.55	340	6
11	-1.25	0.54	341	5
12	-1.14	0.52	342	5
13	-1.02	0.51	343	5
14	-0.90	0.48	345	5
15	-0.78	0.46	346	5
16	-0.65	0.43	347	4
17	-0.54	0.40	348	4
18	-0.43	0.37	349	4
19	-0.32	0.34	350	3
20	-0.23	0.32	351	3
21	-0.14	0.30	352	3
22	-0.05	0.28	353	3
23	0.03	0.27	354	3
24	0.11	0.25	355	3
25	0.19	0.25	356	2
26	0.26	0.24	356	2
27	0.33	0.23	357	2
28	0.40	0.23	358	2
29	0.47	0.22	358	2
30	0.54	0.22	359	2
31	0.61	0.22	360	2
32	0.68	0.22	361	2
33	0.75	0.22	361	2
34	0.82	0.21	362	2
35	0.89	0.21	363	2
36	0.97	0.21	363	2
37	1.04	0.22	364	2
38	1.12	0.22	365	2
39	1.20	0.22	366	2
40	1.28	0.22	367	2
41	1.37	0.23	368	2
42	1.46	0.23	368	2
43	1.55	0.24	369	2
44	1.66	0.25	371	3
45	1.78	0.26	372	3
46	1.91	0.28	373	3
47	2.06	0.31	375	3
48	2.24	0.34	376	3
49	2.47	0.39	379	4
50	2.77	0.46	382	5

Grade 7 Form O

Score	EAP	SD	Score	SE
0	-2.17	0.53	332	5
1	-2.12	0.53	332	5
2	-2.06	0.54	333	5
3	-2.00	0.54	333	6
4	-1.94	0.55	334	6
5	-1.87	0.55	335	6
6	-1.79	0.56	335	6
7	-1.71	0.56	336	6
8	-1.62	0.56	337	6
9	-1.52	0.56	338	6
10	-1.42	0.55	339	6
11	-1.30	0.54	340	5
12	-1.19	0.53	342	5
13	-1.06	0.51	343	5
14	-0.94	0.49	344	5
15	-0.82	0.47	345	5
16	-0.70	0.44	347	4
17	-0.58	0.41	348	4
18	-0.47	0.39	349	4
19	-0.36	0.36	350	4
20	-0.26	0.34	351	3
21	-0.16	0.32	352	3
22	-0.07	0.30	353	3
23	0.02	0.29	354	3
24	0.11	0.28	355	3
25	0.19	0.27	356	3
26	0.27	0.26	356	3
27	0.34	0.25	357	3
28	0.42	0.25	358	2
29	0.49	0.24	359	2
30	0.57	0.24	359	2
31	0.64	0.23	360	2
32	0.71	0.23	361	2
33	0.79	0.23	362	2
34	0.86	0.23	362	2
35	0.94	0.23	363	2
36	1.01	0.23	364	2
37	1.09	0.23	365	2
38	1.17	0.23	366	2
39	1.25	0.24	366	2
40	1.34	0.24	367	2
41	1.43	0.25	368	3
42	1.53	0.26	369	3
43	1.63	0.27	370	3
44	1.75	0.28	371	3
45	1.87	0.29	373	3
46	2.01	0.31	374	3
47	2.17	0.34	376	3
48	2.36	0.37	378	4
49	2.59	0.42	380	4
50	2.87	0.47	383	5

Grade 7 Form P

Score	EAP	SD	Score	SE
0	-2.13	0.52	332	5
1	-2.08	0.53	333	5
2	-2.02	0.53	333	5
3	-1.97	0.54	334	5
4	-1.90	0.54	334	5
5	-1.83	0.54	335	6
6	-1.76	0.55	336	6
7	-1.67	0.55	337	6
8	-1.58	0.55	338	6
9	-1.48	0.55	339	6
10	-1.38	0.54	340	5
11	-1.26	0.53	341	5
12	-1.15	0.51	342	5
13	-1.02	0.49	343	5
14	-0.90	0.47	345	5
15	-0.78	0.44	346	5
16	-0.66	0.42	347	4
17	-0.54	0.39	348	4
18	-0.44	0.36	349	4
19	-0.33	0.34	350	3
20	-0.24	0.32	351	3
21	-0.15	0.30	352	3
22	-0.06	0.28	353	3
23	0.03	0.27	354	3
24	0.11	0.26	355	3
25	0.18	0.25	356	3
26	0.26	0.25	356	2
27	0.33	0.24	357	2
28	0.41	0.24	358	2
29	0.48	0.23	359	2
30	0.55	0.23	359	2
31	0.62	0.23	360	2
32	0.69	0.22	361	2
33	0.76	0.22	361	2
34	0.84	0.22	362	2
35	0.91	0.22	363	2
36	0.98	0.22	364	2
37	1.06	0.22	364	2
38	1.14	0.22	365	2
39	1.22	0.22	366	2
40	1.30	0.23	367	2
41	1.39	0.23	368	2
42	1.48	0.24	369	2
43	1.58	0.25	370	3
44	1.69	0.26	371	3
45	1.80	0.27	372	3
46	1.94	0.29	373	3
47	2.09	0.32	375	3
48	2.27	0.35	377	4
49	2.50	0.40	379	4
50	2.80	0.46	382	5

Grade 8 Form K

Score	EAP	SD	Score	SE
0	-2.32	0.54	332	5
1	-2.28	0.54	333	5
2	-2.23	0.54	333	5
3	-2.18	0.55	334	5
4	-2.13	0.55	334	5
5	-2.07	0.55	335	6
6	-2.01	0.56	335	6
7	-1.95	0.56	336	6
8	-1.88	0.56	337	6
9	-1.81	0.56	337	6
10	-1.73	0.56	338	6
11	-1.65	0.56	339	6
12	-1.57	0.56	340	6
13	-1.47	0.56	341	6
14	-1.37	0.55	342	6
15	-1.27	0.54	343	5
16	-1.17	0.53	344	5
17	-1.06	0.52	345	5
18	-0.95	0.50	346	5
19	-0.83	0.48	347	5
20	-0.72	0.46	348	5
21	-0.61	0.44	349	4
22	-0.51	0.41	350	4
23	-0.41	0.39	351	4
24	-0.31	0.37	352	4
25	-0.21	0.35	353	3
26	-0.12	0.33	354	3
27	-0.04	0.31	355	3
28	0.04	0.30	356	3
29	0.12	0.28	357	3
30	0.20	0.27	357	3
31	0.27	0.26	358	3
32	0.34	0.25	359	3
33	0.41	0.24	360	2
34	0.48	0.23	360	2
35	0.54	0.23	361	2
36	0.60	0.22	361	2
37	0.67	0.21	362	2
38	0.73	0.21	363	2
39	0.79	0.21	363	2
40	0.85	0.20	364	2
41	0.91	0.20	364	2
42	0.97	0.19	365	2
43	1.03	0.19	366	2
44	1.09	0.19	366	2
45	1.15	0.19	367	2
46	1.21	0.19	367	2
47	1.27	0.19	368	2
48	1.33	0.19	369	2
49	1.40	0.19	369	2
50	1.47	0.19	370	2
51	1.54	0.20	371	2
52	1.61	0.20	372	2
53	1.70	0.21	372	2
54	1.79	0.22	373	2
55	1.89	0.24	374	2
56	2.00	0.26	375	3
57	2.14	0.29	377	3
58	2.31	0.32	378	3
59	2.52	0.38	381	4
60	2.79	0.44	383	4

Grade 8 Form L

Score	EAP	SD	Score	SE
0	-2.24	0.53	333	5
1	-2.19	0.54	334	5
2	-2.15	0.54	334	5
3	-2.10	0.54	334	5
4	-2.05	0.55	335	5
5	-2.00	0.55	335	5
6	-1.94	0.55	336	6
7	-1.88	0.56	337	6
8	-1.81	0.56	337	6
9	-1.74	0.56	338	6
10	-1.67	0.56	339	6
11	-1.59	0.56	340	6
12	-1.50	0.56	340	6
13	-1.41	0.55	341	6
14	-1.31	0.55	342	5
15	-1.21	0.54	343	5
16	-1.10	0.52	344	5
17	-0.99	0.51	345	5
18	-0.88	0.49	347	5
19	-0.77	0.47	348	5
20	-0.66	0.44	349	4
21	-0.55	0.42	350	4
22	-0.45	0.39	351	4
23	-0.35	0.37	352	4
24	-0.26	0.34	353	3
25	-0.17	0.32	354	3
26	-0.08	0.31	355	3
27	-0.00	0.29	355	3
28	0.07	0.28	356	3
29	0.15	0.26	357	3
30	0.22	0.25	358	3
31	0.29	0.24	358	2
32	0.36	0.24	359	2
33	0.42	0.23	360	2
34	0.49	0.22	360	2
35	0.55	0.22	361	2
36	0.61	0.21	362	2
37	0.67	0.21	362	2
38	0.73	0.20	363	2
39	0.79	0.20	363	2
40	0.85	0.20	364	2
41	0.91	0.19	365	2
42	0.97	0.19	365	2
43	1.03	0.19	366	2
44	1.09	0.19	366	2
45	1.15	0.19	367	2
46	1.21	0.19	368	2
47	1.28	0.19	368	2
48	1.34	0.19	369	2
49	1.41	0.19	369	2
50	1.48	0.20	370	2
51	1.56	0.20	371	2
52	1.64	0.21	372	2
53	1.73	0.22	373	2
54	1.83	0.23	374	2
55	1.94	0.25	375	3
56	2.08	0.28	376	3
57	2.25	0.32	378	3
58	2.46	0.37	380	4
59	2.74	0.44	383	4

Grade 8 Form M

Score	EAP	SD	Score	SE
0	-2.30	0.52	332	5
1	-2.26	0.53	333	5
2	-2.22	0.53	333	5
3	-2.17	0.54	334	5
4	-2.12	0.54	334	5
5	-2.07	0.54	335	5
6	-2.01	0.55	335	5
7	-1.95	0.55	336	6
8	-1.88	0.55	337	6
9	-1.81	0.56	337	6
10	-1.74	0.56	338	6
11	-1.66	0.56	339	6
12	-1.57	0.56	340	6
13	-1.48	0.56	341	6
14	-1.38	0.55	342	6
15	-1.28	0.54	343	5
16	-1.17	0.53	344	5
17	-1.06	0.52	345	5
18	-0.95	0.50	346	5
19	-0.84	0.48	347	5
20	-0.73	0.46	348	5
21	-0.62	0.44	349	4
22	-0.51	0.42	350	4
23	-0.40	0.39	351	4
24	-0.30	0.37	352	4
25	-0.21	0.35	353	3
26	-0.12	0.33	354	3
27	-0.03	0.31	355	3
28	0.05	0.30	356	3
29	0.13	0.29	357	3
30	0.21	0.27	358	3
31	0.29	0.26	358	3
32	0.36	0.25	359	3
33	0.43	0.25	360	2
34	0.50	0.24	360	2
35	0.57	0.23	361	2
36	0.63	0.23	362	2
37	0.70	0.22	362	2
38	0.76	0.22	363	2
39	0.82	0.21	364	2
40	0.89	0.21	364	2
41	0.95	0.21	365	2
42	1.02	0.20	366	2
43	1.08	0.20	366	2
44	1.14	0.20	367	2
45	1.21	0.20	368	2
46	1.28	0.20	368	2
47	1.35	0.20	369	2
48	1.42	0.21	370	2
49	1.49	0.21	370	2
50	1.57	0.21	371	2
51	1.66	0.22	372	2
52	1.75	0.23	373	2
53	1.85	0.24	374	2
54	1.96	0.26	375	3
55	2.08	0.27	376	3
56	2.23	0.30	378	3
57	2.40	0.34	379	3
58	2.62	0.38	382	4
59	2.90	0.44	384	4

Grade 8 Form N

Score	EAP	SD	Score	SE
0	-2.26	0.53	333	5
1	-2.22	0.53	333	5
2	-2.18	0.53	334	5
3	-2.13	0.54	334	5
4	-2.08	0.54	335	5
5	-2.03	0.55	335	5
6	-1.97	0.55	336	5
7	-1.91	0.55	336	6
8	-1.85	0.56	337	6
9	-1.78	0.56	338	6
10	-1.71	0.56	338	6
11	-1.63	0.56	339	6
12	-1.54	0.56	340	6
13	-1.45	0.56	341	6
14	-1.35	0.55	342	6
15	-1.25	0.54	343	5
16	-1.15	0.53	344	5
17	-1.04	0.52	345	5
18	-0.93	0.50	346	5
19	-0.82	0.48	347	5
20	-0.70	0.46	348	5
21	-0.59	0.43	349	4
22	-0.49	0.41	351	4
23	-0.39	0.39	352	4
24	-0.29	0.36	353	4
25	-0.20	0.34	353	3
26	-0.11	0.32	354	3
27	-0.02	0.31	355	3
28	0.06	0.29	356	3
29	0.14	0.28	357	3
30	0.21	0.27	358	3
31	0.28	0.26	358	3
32	0.35	0.25	359	2
33	0.42	0.24	360	2
34	0.49	0.23	360	2
35	0.56	0.23	361	2
36	0.62	0.22	362	2
37	0.68	0.22	362	2
38	0.75	0.21	363	2
39	0.81	0.21	364	2
40	0.87	0.21	364	2
41	0.94	0.21	365	2
42	1.00	0.20	365	2
43	1.06	0.20	366	2
44	1.13	0.20	367	2
45	1.19	0.20	367	2
46	1.26	0.20	368	2
47	1.33	0.21	369	2
48	1.40	0.21	369	2
49	1.48	0.21	370	2
50	1.56	0.22	371	2
51	1.64	0.22	372	2
52	1.74	0.23	373	2
53	1.84	0.24	374	2
54	1.95	0.26	375	3
55	2.07	0.28	376	3
56	2.22	0.30	378	3
57	2.40	0.34	379	3
58	2.62	0.39	382	4
59	2.90	0.45	384	4

Algebra I Form F

Score	EAP	SD	Score	SE
0	-3.13	0.48	119	5
1	-3.08	0.49	119	5
2	-3.02	0.49	120	5
3	-2.96	0.50	120	5
4	-2.89	0.50	121	5
5	-2.82	0.50	122	5
6	-2.75	0.50	122	5
7	-2.68	0.50	123	5
8	-2.60	0.50	124	5
9	-2.51	0.50	125	5
10	-2.43	0.50	126	5
11	-2.34	0.49	127	5
12	-2.25	0.49	128	5
13	-2.15	0.48	129	5
14	-2.05	0.47	129	5
15	-1.95	0.46	130	5
16	-1.85	0.45	132	4
17	-1.75	0.44	133	4
18	-1.65	0.42	134	4
19	-1.55	0.41	135	4
20	-1.45	0.40	136	4
21	-1.35	0.38	137	4
22	-1.25	0.37	137	4
23	-1.16	0.36	138	4
24	-1.07	0.35	139	3
25	-0.98	0.34	140	3
26	-0.89	0.33	141	3
27	-0.81	0.32	142	3
28	-0.72	0.32	143	3
29	-0.64	0.31	144	3
30	-0.56	0.30	144	3
31	-0.48	0.30	145	3
32	-0.40	0.30	146	3
33	-0.32	0.29	147	3
34	-0.24	0.29	148	3
35	-0.16	0.29	148	3
36	-0.08	0.28	149	3
37	-0.00	0.28	150	3
38	0.07	0.28	151	3
39	0.15	0.28	152	3
40	0.23	0.27	152	3
41	0.31	0.27	153	3
42	0.38	0.27	154	3
43	0.46	0.27	155	3
44	0.54	0.27	155	3
45	0.62	0.27	156	3
46	0.70	0.26	157	3
47	0.78	0.26	158	3
48	0.86	0.26	159	3
49	0.95	0.26	159	3
50	1.03	0.26	160	3
51	1.12	0.26	161	3
52	1.21	0.26	162	3
53	1.30	0.26	163	3
54	1.39	0.27	164	3
55	1.49	0.27	165	3
56	1.60	0.28	166	3
57	1.71	0.29	167	3
58	1.83	0.30	168	3
59	1.96	0.32	170	3
60	2.10	0.34	171	3
61	2.27	0.37	173	4
62	2.45	0.40	175	4
63	2.67	0.44	177	4
64	2.91	0.48	179	5

Algebra I Form G

Score	EAP	SD	Score	SE
0	-3.10	0.47	119	5
1	-3.05	0.47	119	5
2	-3.00	0.47	120	5
3	-2.94	0.48	121	5
4	-2.88	0.48	121	5
5	-2.82	0.48	122	5
6	-2.75	0.48	122	5
7	-2.68	0.48	123	5
8	-2.60	0.49	124	5
9	-2.52	0.48	125	5
10	-2.43	0.48	126	5
11	-2.34	0.48	127	5
12	-2.25	0.47	128	5
13	-2.15	0.47	129	5
14	-2.05	0.46	130	5
15	-1.95	0.44	131	4
16	-1.84	0.43	132	4
17	-1.74	0.42	133	4
18	-1.64	0.41	134	4
19	-1.54	0.39	135	4
20	-1.44	0.38	136	4
21	-1.35	0.37	137	4
22	-1.25	0.36	137	4
23	-1.16	0.35	138	3
24	-1.07	0.34	139	3
25	-0.99	0.33	140	3
26	-0.90	0.32	141	3
27	-0.82	0.31	142	3
28	-0.73	0.31	143	3
29	-0.65	0.30	143	3
30	-0.57	0.30	144	3
31	-0.49	0.30	145	3
32	-0.41	0.29	146	3
33	-0.33	0.29	147	3
34	-0.25	0.29	147	3
35	-0.18	0.28	148	3
36	-0.10	0.28	149	3
37	-0.02	0.28	150	3
38	0.06	0.28	151	3
39	0.14	0.28	151	3
40	0.22	0.28	152	3
41	0.30	0.28	153	3
42	0.38	0.28	154	3
43	0.46	0.28	155	3
44	0.54	0.28	155	3
45	0.62	0.28	156	3
46	0.70	0.28	157	3
47	0.79	0.28	158	3
48	0.87	0.28	159	3
49	0.96	0.28	160	3
50	1.05	0.28	161	3
51	1.14	0.28	161	3
52	1.23	0.28	162	3
53	1.33	0.28	163	3
54	1.43	0.29	164	3
55	1.53	0.29	165	3
56	1.64	0.30	166	3
57	1.76	0.30	168	3
58	1.88	0.31	169	3
59	2.02	0.33	170	3
60	2.16	0.34	172	3
61	2.33	0.37	173	4
62	2.52	0.39	175	4
63	2.74	0.43	177	4
64	3.00	0.47	180	5

Algebra I Form H

Score	EAP	SD	Score	SE
0	-3.20	0.46	118	5
1	-3.15	0.47	118	5
2	-3.10	0.47	119	5
3	-3.04	0.48	120	5
4	-2.98	0.48	120	5
5	-2.92	0.48	121	5
6	-2.85	0.48	121	5
7	-2.78	0.49	122	5
8	-2.70	0.49	123	5
9	-2.63	0.49	124	5
10	-2.54	0.48	125	5
11	-2.45	0.48	125	5
12	-2.36	0.48	126	5
13	-2.27	0.47	127	5
14	-2.17	0.46	128	5
15	-2.07	0.45	129	5
16	-1.97	0.44	130	4
17	-1.87	0.43	131	4
18	-1.77	0.42	132	4
19	-1.67	0.40	133	4
20	-1.57	0.39	134	4
21	-1.47	0.38	135	4
22	-1.38	0.37	136	4
23	-1.29	0.36	137	4
24	-1.20	0.35	138	3
25	-1.11	0.34	139	3
26	-1.02	0.33	140	3
27	-0.94	0.32	141	3
28	-0.85	0.32	141	3
29	-0.77	0.31	142	3
30	-0.69	0.31	143	3
31	-0.61	0.30	144	3
32	-0.53	0.30	145	3
33	-0.45	0.29	145	3
34	-0.37	0.29	146	3
35	-0.30	0.29	147	3
36	-0.22	0.28	148	3
37	-0.14	0.28	149	3
38	-0.06	0.28	149	3
39	0.01	0.28	150	3
40	0.09	0.27	151	3
41	0.17	0.27	152	3
42	0.25	0.27	152	3
43	0.33	0.27	153	3
44	0.41	0.27	154	3
45	0.48	0.27	155	3
46	0.56	0.27	156	3
47	0.65	0.27	156	3
48	0.73	0.26	157	3
49	0.81	0.26	158	3
50	0.90	0.27	159	3
51	0.99	0.27	160	3
52	1.08	0.27	161	3
53	1.17	0.27	162	3
54	1.27	0.27	163	3
55	1.37	0.28	164	3
56	1.47	0.29	165	3
57	1.59	0.29	166	3
58	1.71	0.30	167	3
59	1.84	0.32	168	3
60	1.99	0.34	170	3
61	2.16	0.36	172	4
62	2.35	0.39	173	4
63	2.58	0.43	176	4
64	2.85	0.48	178	5

Algebra I Form I

Score	EAP	SD	Score	SE
0	-3.17	0.45	118	5
1	-3.12	0.46	119	5
2	-3.06	0.46	119	5
3	-3.00	0.46	120	5
4	-2.93	0.47	121	5
5	-2.86	0.47	121	5
6	-2.78	0.47	122	5
7	-2.70	0.47	123	5
8	-2.61	0.47	124	5
9	-2.52	0.46	125	5
10	-2.42	0.46	126	5
11	-2.32	0.45	127	4
12	-2.22	0.44	128	4
13	-2.12	0.43	129	4
14	-2.02	0.41	130	4
15	-1.92	0.40	131	4
16	-1.82	0.38	132	4
17	-1.72	0.37	133	4
18	-1.62	0.36	134	4
19	-1.53	0.34	135	3
20	-1.44	0.33	136	3
21	-1.36	0.32	136	3
22	-1.27	0.31	137	3
23	-1.19	0.31	138	3
24	-1.11	0.30	139	3
25	-1.03	0.29	140	3
26	-0.95	0.29	141	3
27	-0.87	0.28	141	3
28	-0.79	0.28	142	3
29	-0.71	0.28	143	3
30	-0.64	0.28	144	3
31	-0.56	0.28	144	3
32	-0.49	0.28	145	3
33	-0.41	0.28	146	3
34	-0.33	0.28	147	3
35	-0.26	0.28	147	3
36	-0.18	0.28	148	3
37	-0.10	0.28	149	3
38	-0.02	0.28	150	3
39	0.06	0.28	151	3
40	0.14	0.28	151	3
41	0.22	0.28	152	3
42	0.30	0.29	153	3
43	0.38	0.29	154	3
44	0.47	0.29	155	3
45	0.56	0.29	156	3
46	0.65	0.29	156	3
47	0.74	0.30	157	3
48	0.83	0.30	158	3
49	0.92	0.30	159	3
50	1.02	0.30	160	3
51	1.12	0.31	161	3
52	1.22	0.31	162	3
53	1.33	0.32	163	3
54	1.44	0.32	164	3
55	1.56	0.33	166	3
56	1.68	0.34	167	3
57	1.81	0.35	168	3
58	1.95	0.36	169	4
59	2.10	0.37	171	4
60	2.26	0.39	173	4
61	2.43	0.41	174	4
62	2.63	0.44	176	4
63	2.85	0.47	178	5
64	3.09	0.50	181	5

Algebra I Form J

Score	EAP	SD	Score	SE
0	-3.21	0.47	118	5
1	-3.15	0.48	118	5
2	-3.09	0.48	119	5
3	-3.03	0.49	120	5
4	-2.96	0.49	120	5
5	-2.89	0.50	121	5
6	-2.82	0.50	122	5
7	-2.74	0.50	123	5
8	-2.66	0.50	123	5
9	-2.57	0.50	124	5
10	-2.48	0.49	125	5
11	-2.39	0.49	126	5
12	-2.29	0.49	127	5
13	-2.19	0.48	128	5
14	-2.09	0.47	129	5
15	-1.99	0.46	130	5
16	-1.89	0.45	131	4
17	-1.78	0.43	132	4
18	-1.68	0.42	133	4
19	-1.58	0.41	134	4
20	-1.48	0.39	135	4
21	-1.38	0.38	136	4
22	-1.28	0.37	137	4
23	-1.19	0.36	138	4
24	-1.10	0.35	139	3
25	-1.01	0.34	140	3
26	-0.92	0.33	141	3
27	-0.83	0.32	142	3
28	-0.75	0.32	143	3
29	-0.66	0.31	143	3
30	-0.58	0.30	144	3
31	-0.50	0.30	145	3
32	-0.41	0.30	146	3
33	-0.33	0.29	147	3
34	-0.25	0.29	147	3
35	-0.17	0.29	148	3
36	-0.09	0.28	149	3
37	-0.01	0.28	150	3
38	0.07	0.28	151	3
39	0.15	0.28	151	3
40	0.23	0.28	152	3
41	0.31	0.28	153	3
42	0.39	0.27	154	3
43	0.47	0.27	155	3
44	0.55	0.27	156	3
45	0.63	0.27	156	3
46	0.72	0.27	157	3
47	0.80	0.27	158	3
48	0.89	0.27	159	3
49	0.97	0.27	160	3
50	1.06	0.27	161	3
51	1.15	0.26	161	3
52	1.24	0.26	162	3
53	1.33	0.26	163	3
54	1.43	0.26	164	3
55	1.53	0.27	165	3
56	1.64	0.27	166	3
57	1.75	0.28	167	3
58	1.87	0.29	169	3
59	2.00	0.30	170	3
60	2.14	0.32	171	3
61	2.30	0.35	173	3
62	2.48	0.38	175	4
63	2.69	0.43	177	4
64	2.93	0.47	179	5

Geometry Form F

Score	EAP	SD	Score	SE
0	-2.64	0.51	124	5
1	-2.59	0.51	124	5
2	-2.53	0.52	125	5
3	-2.48	0.52	125	5
4	-2.41	0.53	126	5
5	-2.35	0.53	127	5
6	-2.28	0.53	127	5
7	-2.21	0.53	128	5
8	-2.13	0.53	129	5
9	-2.04	0.53	130	5
10	-1.96	0.53	130	5
11	-1.86	0.53	131	5
12	-1.77	0.52	132	5
13	-1.67	0.52	133	5
14	-1.56	0.51	134	5
15	-1.46	0.49	135	5
16	-1.35	0.48	137	5
17	-1.24	0.46	138	5
18	-1.13	0.44	139	4
19	-1.02	0.42	140	4
20	-0.92	0.40	141	4
21	-0.82	0.38	142	4
22	-0.72	0.36	143	4
23	-0.63	0.35	144	3
24	-0.54	0.33	145	3
25	-0.45	0.32	146	3
26	-0.36	0.30	146	3
27	-0.28	0.29	147	3
28	-0.20	0.28	148	3
29	-0.13	0.27	149	3
30	-0.05	0.26	149	3
31	0.02	0.26	150	3
32	0.09	0.25	151	3
33	0.16	0.25	152	2
34	0.23	0.24	152	2
35	0.29	0.24	153	2
36	0.36	0.23	154	2
37	0.43	0.23	154	2
38	0.50	0.23	155	2
39	0.56	0.23	156	2
40	0.63	0.23	156	2
41	0.70	0.22	157	2
42	0.76	0.22	158	2
43	0.83	0.22	158	2
44	0.90	0.22	159	2
45	0.97	0.22	160	2
46	1.04	0.22	160	2
47	1.12	0.23	161	2
48	1.19	0.23	162	2
49	1.27	0.23	163	2
50	1.36	0.24	164	2
51	1.44	0.24	164	2
52	1.54	0.25	165	3
53	1.64	0.26	166	3
54	1.75	0.27	167	3
55	1.87	0.29	169	3
56	2.00	0.31	170	3
57	2.16	0.34	172	3
58	2.35	0.37	173	4
59	2.57	0.42	176	4
60	2.84	0.47	178	5

Geometry Form G

Score	EAP	SD	Score	SE
0	-2.69	0.50	123	5
1	-2.64	0.50	124	5
2	-2.59	0.51	124	5
3	-2.54	0.51	125	5
4	-2.49	0.51	125	5
5	-2.43	0.52	126	5
6	-2.36	0.52	126	5
7	-2.30	0.52	127	5
8	-2.22	0.52	128	5
9	-2.14	0.52	129	5
10	-2.06	0.52	129	5
11	-1.97	0.52	130	5
12	-1.88	0.51	131	5
13	-1.78	0.51	132	5
14	-1.68	0.50	133	5
15	-1.57	0.49	134	5
16	-1.46	0.47	135	5
17	-1.35	0.46	136	5
18	-1.24	0.44	138	4
19	-1.14	0.42	139	4
20	-1.03	0.40	140	4
21	-0.93	0.39	141	4
22	-0.83	0.37	142	4
23	-0.74	0.36	143	4
24	-0.64	0.34	144	3
25	-0.55	0.33	144	3
26	-0.47	0.32	145	3
27	-0.38	0.31	146	3
28	-0.30	0.30	147	3
29	-0.22	0.29	148	3
30	-0.14	0.29	149	3
31	-0.06	0.28	149	3
32	0.02	0.27	150	3
33	0.09	0.27	151	3
34	0.16	0.26	152	3
35	0.24	0.26	152	3
36	0.31	0.25	153	3
37	0.38	0.25	154	2
38	0.45	0.25	155	2
39	0.53	0.24	155	2
40	0.60	0.24	156	2
41	0.67	0.24	157	2
42	0.74	0.24	157	2
43	0.81	0.24	158	2
44	0.89	0.23	159	2
45	0.96	0.23	160	2
46	1.04	0.23	160	2
47	1.11	0.24	161	2
48	1.19	0.24	162	2
49	1.28	0.24	163	2
50	1.36	0.24	164	2
51	1.45	0.25	165	3
52	1.55	0.26	165	3
53	1.65	0.27	167	3
54	1.77	0.28	168	3
55	1.89	0.30	169	3
56	2.03	0.32	170	3
57	2.19	0.35	172	3
58	2.37	0.38	174	4
59	2.59	0.42	176	4
60	2.86	0.48	179	5

Geometry Form H

Score	EAP	SD	Score	SE
0	-2.45	0.51	126	5
1	-2.40	0.52	126	5
2	-2.36	0.52	126	5
3	-2.31	0.52	127	5
4	-2.26	0.53	127	5
5	-2.21	0.53	128	5
6	-2.15	0.53	129	5
7	-2.09	0.53	129	5
8	-2.02	0.53	130	5
9	-1.95	0.53	130	5
10	-1.88	0.53	131	5
11	-1.80	0.53	132	5
12	-1.72	0.53	133	5
13	-1.63	0.52	134	5
14	-1.53	0.52	135	5
15	-1.43	0.51	136	5
16	-1.33	0.49	137	5
17	-1.23	0.48	138	5
18	-1.12	0.46	139	5
19	-1.01	0.44	140	4
20	-0.91	0.42	141	4
21	-0.81	0.39	142	4
22	-0.71	0.37	143	4
23	-0.61	0.35	144	4
24	-0.52	0.33	145	3
25	-0.44	0.31	146	3
26	-0.35	0.30	146	3
27	-0.27	0.29	147	3
28	-0.20	0.28	148	3
29	-0.13	0.27	149	3
30	-0.05	0.26	149	3
31	0.02	0.25	150	3
32	0.08	0.25	151	2
33	0.15	0.24	152	2
34	0.22	0.24	152	2
35	0.28	0.23	153	2
36	0.35	0.23	153	2
37	0.41	0.23	154	2
38	0.47	0.23	155	2
39	0.54	0.22	155	2
40	0.60	0.22	156	2
41	0.67	0.22	157	2
42	0.73	0.22	157	2
43	0.80	0.22	158	2
44	0.87	0.22	159	2
45	0.94	0.22	159	2
46	1.01	0.23	160	2
47	1.08	0.23	161	2
48	1.16	0.23	162	2
49	1.24	0.24	162	2
50	1.32	0.24	163	2
51	1.41	0.25	164	2
52	1.50	0.26	165	3
53	1.60	0.27	166	3
54	1.71	0.28	167	3
55	1.83	0.29	168	3
56	1.97	0.31	170	3
57	2.13	0.34	171	3
58	2.31	0.37	173	4
59	2.54	0.42	175	4
60	2.82	0.47	178	5

Geometry Form I

Score	EAP	SD	Score	SE
0	-2.55	0.52	124	5
1	-2.51	0.52	125	5
2	-2.46	0.52	125	5
3	-2.40	0.53	126	5
4	-2.35	0.53	127	5
5	-2.29	0.53	127	5
6	-2.23	0.53	128	5
7	-2.16	0.53	128	5
8	-2.09	0.53	129	5
9	-2.01	0.53	130	5
10	-1.93	0.53	131	5
11	-1.84	0.53	132	5
12	-1.75	0.52	132	5
13	-1.66	0.52	133	5
14	-1.56	0.51	134	5
15	-1.46	0.50	135	5
16	-1.35	0.48	136	5
17	-1.24	0.46	138	5
18	-1.14	0.44	139	4
19	-1.03	0.42	140	4
20	-0.93	0.40	141	4
21	-0.83	0.38	142	4
22	-0.73	0.36	143	4
23	-0.64	0.35	144	3
24	-0.55	0.33	144	3
25	-0.47	0.31	145	3
26	-0.39	0.30	146	3
27	-0.31	0.29	147	3
28	-0.23	0.28	148	3
29	-0.16	0.27	148	3
30	-0.09	0.26	149	3
31	-0.02	0.25	150	3
32	0.05	0.25	150	2
33	0.12	0.24	151	2
34	0.18	0.24	152	2
35	0.25	0.23	152	2
36	0.31	0.23	153	2
37	0.37	0.22	154	2
38	0.44	0.22	154	2
39	0.50	0.22	155	2
40	0.56	0.22	156	2
41	0.62	0.21	156	2
42	0.69	0.21	157	2
43	0.75	0.21	158	2
44	0.82	0.21	158	2
45	0.88	0.22	159	2
46	0.95	0.22	160	2
47	1.02	0.22	160	2
48	1.10	0.22	161	2
49	1.17	0.23	162	2
50	1.25	0.23	163	2
51	1.34	0.24	163	2
52	1.43	0.25	164	3
53	1.53	0.26	165	3
54	1.64	0.28	166	3
55	1.76	0.29	168	3
56	1.90	0.32	169	3
57	2.06	0.35	171	3
58	2.24	0.38	172	4
59	2.47	0.43	175	4
60	2.74	0.48	177	5

Algebra II Form F

Score	EAP	SD	Score	SE
0	-2.45	0.50	126	5
1	-2.40	0.50	126	5
2	-2.36	0.51	126	5
3	-2.31	0.51	127	5
4	-2.26	0.51	127	5
5	-2.20	0.52	128	5
6	-2.14	0.52	129	5
7	-2.08	0.52	129	5
8	-2.01	0.52	130	5
9	-1.93	0.52	131	5
10	-1.85	0.52	131	5
11	-1.76	0.51	132	5
12	-1.67	0.51	133	5
13	-1.58	0.50	134	5
14	-1.47	0.49	135	5
15	-1.37	0.47	136	5
16	-1.27	0.45	137	5
17	-1.16	0.43	138	4
18	-1.06	0.41	139	4
19	-0.96	0.39	140	4
20	-0.86	0.36	141	4
21	-0.77	0.34	142	3
22	-0.68	0.33	143	3
23	-0.60	0.31	144	3
24	-0.52	0.29	145	3
25	-0.44	0.28	146	3
26	-0.37	0.27	146	3
27	-0.30	0.26	147	3
28	-0.23	0.25	148	3
29	-0.17	0.24	148	2
30	-0.11	0.24	149	2
31	-0.04	0.23	150	2
32	0.02	0.22	150	2
33	0.07	0.22	151	2
34	0.13	0.21	151	2
35	0.19	0.21	152	2
36	0.24	0.21	152	2
37	0.30	0.20	153	2
38	0.35	0.20	154	2
39	0.41	0.20	154	2
40	0.46	0.20	155	2
41	0.52	0.19	155	2
42	0.57	0.19	156	2
43	0.62	0.19	156	2
44	0.68	0.19	157	2
45	0.73	0.19	157	2
46	0.79	0.19	158	2
47	0.85	0.19	158	2
48	0.90	0.19	159	2
49	0.96	0.19	160	2
50	1.02	0.20	160	2
51	1.09	0.20	161	2
52	1.15	0.20	162	2
53	1.22	0.21	162	2
54	1.29	0.21	163	2
55	1.37	0.22	164	2
56	1.45	0.23	165	2
57	1.54	0.24	165	2
58	1.64	0.25	166	2
59	1.75	0.26	167	3
60	1.87	0.28	169	3
61	2.02	0.31	170	3
62	2.19	0.35	172	3
63	2.41	0.40	174	4
64	2.71	0.46	177	5

Algebra II Form G

Score	EAP	SD	Score	SE
0	-2.52	0.51	125	5
1	-2.48	0.52	125	5
2	-2.43	0.52	126	5
3	-2.38	0.52	126	5
4	-2.33	0.53	127	5
5	-2.28	0.53	127	5
6	-2.22	0.53	128	5
7	-2.16	0.53	128	5
8	-2.09	0.53	129	5
9	-2.02	0.53	130	5
10	-1.95	0.53	130	5
11	-1.87	0.53	131	5
12	-1.79	0.53	132	5
13	-1.70	0.53	133	5
14	-1.61	0.52	134	5
15	-1.51	0.51	135	5
16	-1.41	0.50	136	5
17	-1.30	0.48	137	5
18	-1.20	0.47	138	5
19	-1.10	0.45	139	4
20	-0.99	0.43	140	4
21	-0.89	0.41	141	4
22	-0.79	0.38	142	4
23	-0.70	0.36	143	4
24	-0.61	0.35	144	3
25	-0.52	0.33	145	3
26	-0.44	0.31	146	3
27	-0.36	0.30	146	3
28	-0.28	0.29	147	3
29	-0.21	0.27	148	3
30	-0.14	0.26	149	3
31	-0.07	0.26	149	3
32	-0.00	0.25	150	2
33	0.06	0.24	151	2
34	0.13	0.23	151	2
35	0.19	0.23	152	2
36	0.25	0.22	152	2
37	0.31	0.22	153	2
38	0.37	0.22	154	2
39	0.43	0.21	154	2
40	0.48	0.21	155	2
41	0.54	0.21	155	2
42	0.60	0.20	156	2
43	0.66	0.20	157	2
44	0.72	0.20	157	2
45	0.78	0.20	158	2
46	0.83	0.20	158	2
47	0.89	0.20	159	2
48	0.96	0.20	160	2
49	1.02	0.20	160	2
50	1.08	0.20	161	2
51	1.15	0.21	161	2
52	1.22	0.21	162	2
53	1.29	0.21	163	2
54	1.37	0.22	164	2
55	1.45	0.23	164	2
56	1.53	0.23	165	2
57	1.63	0.24	166	2
58	1.73	0.26	167	3
59	1.84	0.27	168	3
60	1.97	0.29	170	3
61	2.12	0.32	171	3
62	2.30	0.35	173	4
63	2.52	0.40	175	4
64	2.81	0.46	178	5

Algebra II Form H

Score	EAP	SD	Score	SE
0	-2.47	0.53	125	5
1	-2.42	0.53	126	5
2	-2.37	0.53	126	5
3	-2.32	0.53	127	5
4	-2.26	0.53	127	5
5	-2.21	0.54	128	5
6	-2.15	0.54	129	5
7	-2.09	0.54	129	5
8	-2.02	0.54	130	5
9	-1.95	0.54	130	5
10	-1.88	0.54	131	5
11	-1.81	0.54	132	5
12	-1.72	0.54	133	5
13	-1.64	0.53	134	5
14	-1.55	0.53	135	5
15	-1.45	0.52	136	5
16	-1.35	0.51	137	5
17	-1.25	0.49	138	5
18	-1.14	0.47	139	5
19	-1.04	0.45	140	5
20	-0.93	0.43	141	4
21	-0.83	0.41	142	4
22	-0.73	0.39	143	4
23	-0.63	0.37	144	4
24	-0.54	0.34	145	3
25	-0.46	0.33	145	3
26	-0.37	0.31	146	3
27	-0.29	0.29	147	3
28	-0.22	0.28	148	3
29	-0.15	0.27	149	3
30	-0.08	0.26	149	3
31	-0.01	0.25	150	2
32	0.06	0.24	151	2
33	0.12	0.23	151	2
34	0.18	0.23	152	2
35	0.24	0.22	152	2
36	0.30	0.22	153	2
37	0.36	0.21	154	2
38	0.42	0.21	154	2
39	0.48	0.21	155	2
40	0.53	0.21	155	2
41	0.59	0.20	156	2
42	0.65	0.20	156	2
43	0.71	0.20	157	2
44	0.76	0.20	158	2
45	0.82	0.20	158	2
46	0.88	0.20	159	2
47	0.94	0.20	159	2
48	1.00	0.20	160	2
49	1.07	0.21	161	2
50	1.13	0.21	161	2
51	1.20	0.21	162	2
52	1.27	0.22	163	2
53	1.34	0.22	163	2
54	1.42	0.23	164	2
55	1.51	0.24	165	2
56	1.60	0.25	166	2
57	1.70	0.26	167	3
58	1.81	0.28	168	3
59	1.93	0.29	169	3
60	2.07	0.32	171	3
61	2.23	0.34	172	3
62	2.41	0.38	174	4
63	2.64	0.42	176	4
64	2.91	0.48	179	5

Algebra II Form I

Score	EAP	SD	Score	SE
0	-2.49	0.50	125	5
1	-2.45	0.50	125	5
2	-2.41	0.50	126	5
3	-2.36	0.51	126	5
4	-2.31	0.51	127	5
5	-2.25	0.51	128	5
6	-2.19	0.51	128	5
7	-2.12	0.51	129	5
8	-2.05	0.51	129	5
9	-1.98	0.51	130	5
10	-1.90	0.51	131	5
11	-1.81	0.51	132	5
12	-1.72	0.50	133	5
13	-1.63	0.49	134	5
14	-1.53	0.48	135	5
15	-1.42	0.47	136	5
16	-1.32	0.45	137	4
17	-1.22	0.43	138	4
18	-1.11	0.41	139	4
19	-1.01	0.38	140	4
20	-0.92	0.36	141	4
21	-0.83	0.34	142	3
22	-0.74	0.32	143	3
23	-0.66	0.31	143	3
24	-0.58	0.29	144	3
25	-0.50	0.28	145	3
26	-0.43	0.27	146	3
27	-0.36	0.26	146	3
28	-0.30	0.25	147	2
29	-0.23	0.24	148	2
30	-0.17	0.24	148	2
31	-0.11	0.23	149	2
32	-0.05	0.22	150	2
33	0.01	0.22	150	2
34	0.07	0.21	151	2
35	0.13	0.21	151	2
36	0.18	0.21	152	2
37	0.24	0.20	152	2
38	0.29	0.20	153	2
39	0.35	0.20	153	2
40	0.40	0.20	154	2
41	0.46	0.19	155	2
42	0.51	0.19	155	2
43	0.57	0.19	156	2
44	0.62	0.19	156	2
45	0.68	0.19	157	2
46	0.73	0.19	157	2
47	0.79	0.19	158	2
48	0.85	0.19	158	2
49	0.90	0.19	159	2
50	0.96	0.19	160	2
51	1.03	0.20	160	2
52	1.09	0.20	161	2
53	1.16	0.20	162	2
54	1.23	0.21	162	2
55	1.30	0.21	163	2
56	1.39	0.22	164	2
57	1.47	0.23	165	2
58	1.57	0.24	166	2
59	1.68	0.26	167	3
60	1.80	0.28	168	3
61	1.94	0.30	169	3
62	2.12	0.34	171	3
63	2.34	0.39	173	4
64	2.64	0.46	176	5

Algebra II Form J

Score	EAP	SD	Score	SE
0	-2.53	0.51	125	5
1	-2.48	0.52	125	5
2	-2.44	0.52	126	5
3	-2.39	0.52	126	5
4	-2.33	0.53	127	5
5	-2.27	0.53	127	5
6	-2.21	0.53	128	5
7	-2.15	0.53	129	5
8	-2.08	0.53	129	5
9	-2.00	0.53	130	5
10	-1.92	0.53	131	5
11	-1.84	0.53	132	5
12	-1.75	0.53	132	5
13	-1.66	0.52	133	5
14	-1.56	0.51	134	5
15	-1.46	0.50	135	5
16	-1.35	0.48	136	5
17	-1.25	0.47	137	5
18	-1.15	0.45	139	4
19	-1.04	0.43	140	4
20	-0.94	0.41	141	4
21	-0.84	0.39	142	4
22	-0.75	0.37	143	4
23	-0.66	0.35	143	3
24	-0.57	0.33	144	3
25	-0.48	0.31	145	3
26	-0.40	0.30	146	3
27	-0.33	0.29	147	3
28	-0.25	0.28	147	3
29	-0.18	0.27	148	3
30	-0.11	0.26	149	3
31	-0.05	0.25	150	2
32	0.02	0.24	150	2
33	0.08	0.23	151	2
34	0.14	0.23	151	2
35	0.20	0.22	152	2
36	0.26	0.22	153	2
37	0.32	0.22	153	2
38	0.38	0.21	154	2
39	0.44	0.21	154	2
40	0.49	0.21	155	2
41	0.55	0.20	156	2
42	0.61	0.20	156	2
43	0.67	0.20	157	2
44	0.72	0.20	157	2
45	0.78	0.20	158	2
46	0.84	0.20	158	2
47	0.90	0.20	159	2
48	0.96	0.20	160	2
49	1.03	0.20	160	2
50	1.09	0.21	161	2
51	1.16	0.21	162	2
52	1.23	0.21	162	2
53	1.30	0.22	163	2
54	1.37	0.22	164	2
55	1.46	0.23	165	2
56	1.54	0.24	165	2
57	1.64	0.25	166	2
58	1.74	0.26	167	3
59	1.85	0.27	169	3
60	1.98	0.29	170	3
61	2.14	0.32	171	3
62	2.32	0.35	173	4
63	2.54	0.40	175	4
64	2.83	0.46	178	5

Algebra II Form K

Score	EAP	SD	Score	SE
0	-2.30	0.53	127	5
1	-2.26	0.53	127	5
2	-2.21	0.53	128	5
3	-2.17	0.53	128	5
4	-2.11	0.54	129	5
5	-2.06	0.54	129	5
6	-2.00	0.54	130	5
7	-1.94	0.55	131	5
8	-1.87	0.55	131	5
9	-1.79	0.55	132	5
10	-1.71	0.55	133	5
11	-1.63	0.55	134	5
12	-1.53	0.54	135	5
13	-1.43	0.54	136	5
14	-1.33	0.53	137	5
15	-1.21	0.52	138	5
16	-1.10	0.50	139	5
17	-0.98	0.48	140	5
18	-0.86	0.46	141	5
19	-0.75	0.44	143	4
20	-0.63	0.41	144	4
21	-0.52	0.39	145	4
22	-0.42	0.37	146	4
23	-0.32	0.35	147	3
24	-0.22	0.33	148	3
25	-0.13	0.32	149	3
26	-0.04	0.30	150	3
27	0.05	0.29	150	3
28	0.13	0.28	151	3
29	0.21	0.27	152	3
30	0.29	0.27	153	3
31	0.37	0.26	154	3
32	0.45	0.26	154	3
33	0.53	0.25	155	3
34	0.60	0.25	156	2
35	0.68	0.25	157	2
36	0.76	0.25	158	2
37	0.84	0.25	158	2
38	0.92	0.25	159	2
39	1.00	0.25	160	2
40	1.09	0.25	161	2
41	1.18	0.25	162	3
42	1.27	0.26	163	3
43	1.37	0.27	164	3
44	1.48	0.28	165	3
45	1.60	0.29	166	3
46	1.72	0.30	167	3
47	1.87	0.32	169	3
48	2.03	0.35	170	4
49	2.23	0.39	172	4
50	2.46	0.43	175	4
51	2.74	0.48	177	5

Algebra II Form L

Score	EAP	SD	Score	SE
0	-2.47	0.53	125	5
1	-2.42	0.53	126	5
2	-2.36	0.53	126	5
3	-2.30	0.54	127	5
4	-2.23	0.54	128	5
5	-2.16	0.54	128	5
6	-2.09	0.54	129	5
7	-2.01	0.55	130	5
8	-1.93	0.55	131	5
9	-1.84	0.54	132	5
10	-1.74	0.54	133	5
11	-1.64	0.54	134	5
12	-1.53	0.53	135	5
13	-1.42	0.52	136	5
14	-1.31	0.51	137	5
15	-1.19	0.49	138	5
16	-1.07	0.47	139	5
17	-0.95	0.45	140	4
18	-0.84	0.43	142	4
19	-0.72	0.40	143	4
20	-0.61	0.38	144	4
21	-0.51	0.36	145	4
22	-0.41	0.34	146	3
23	-0.31	0.32	147	3
24	-0.22	0.31	148	3
25	-0.13	0.29	149	3
26	-0.05	0.28	150	3
27	0.03	0.27	150	3
28	0.11	0.26	151	3
29	0.19	0.25	152	3
30	0.27	0.25	153	2
31	0.34	0.24	153	2
32	0.42	0.24	154	2
33	0.49	0.24	155	2
34	0.57	0.23	156	2
35	0.64	0.23	156	2
36	0.72	0.23	157	2
37	0.79	0.23	158	2
38	0.87	0.23	159	2
39	0.95	0.23	160	2
40	1.04	0.24	160	2
41	1.13	0.24	161	2
42	1.22	0.25	162	2
43	1.31	0.25	163	3
44	1.42	0.26	164	3
45	1.53	0.27	165	3
46	1.66	0.29	167	3
47	1.80	0.31	168	3
48	1.96	0.34	170	3
49	2.16	0.37	172	4
50	2.40	0.42	174	4
51	2.69	0.48	177	5

Algebra II Form M

Score	EAP	SD	Score	SE
0	-2.33	0.52	127	5
1	-2.29	0.52	127	5
2	-2.24	0.52	128	5
3	-2.19	0.53	128	5
4	-2.14	0.53	129	5
5	-2.08	0.53	129	5
6	-2.01	0.54	130	5
7	-1.94	0.54	131	5
8	-1.86	0.54	131	5
9	-1.78	0.54	132	5
10	-1.69	0.54	133	5
11	-1.59	0.53	134	5
12	-1.49	0.53	135	5
13	-1.38	0.52	136	5
14	-1.27	0.50	137	5
15	-1.15	0.48	138	5
16	-1.03	0.46	140	5
17	-0.92	0.44	141	4
18	-0.80	0.41	142	4
19	-0.69	0.39	143	4
20	-0.59	0.37	144	4
21	-0.49	0.34	145	3
22	-0.39	0.33	146	3
23	-0.30	0.31	147	3
24	-0.21	0.29	148	3
25	-0.13	0.28	149	3
26	-0.05	0.27	149	3
27	0.03	0.26	150	3
28	0.10	0.25	151	3
29	0.18	0.25	152	2
30	0.25	0.24	152	2
31	0.32	0.23	153	2
32	0.39	0.23	154	2
33	0.46	0.23	155	2
34	0.53	0.22	155	2
35	0.60	0.22	156	2
36	0.67	0.22	157	2
37	0.75	0.22	157	2
38	0.82	0.22	158	2
39	0.89	0.22	159	2
40	0.97	0.23	160	2
41	1.05	0.23	161	2
42	1.14	0.23	161	2
43	1.23	0.24	162	2
44	1.32	0.25	163	3
45	1.43	0.26	164	3
46	1.55	0.28	165	3
47	1.68	0.30	167	3
48	1.84	0.33	168	3
49	2.03	0.37	170	4
50	2.26	0.42	173	4
51	2.55	0.48	176	5

Algebra II Form N

Score	EAP	SD	Score	SE
0	-2.27	0.54	127	5
1	-2.21	0.54	128	5
2	-2.15	0.54	128	5
3	-2.10	0.54	129	5
4	-2.04	0.54	130	5
5	-1.98	0.54	130	5
6	-1.91	0.54	131	5
7	-1.85	0.54	132	5
8	-1.77	0.54	132	5
9	-1.70	0.54	133	5
10	-1.61	0.54	134	5
11	-1.52	0.54	135	5
12	-1.42	0.53	136	5
13	-1.32	0.52	137	5
14	-1.21	0.51	138	5
15	-1.10	0.50	139	5
16	-0.98	0.47	140	5
17	-0.86	0.45	141	4
18	-0.75	0.42	143	4
19	-0.63	0.39	144	4
20	-0.53	0.37	145	4
21	-0.42	0.34	146	3
22	-0.33	0.32	147	3
23	-0.24	0.30	148	3
24	-0.15	0.28	148	3
25	-0.07	0.27	149	3
26	0.01	0.26	150	3
27	0.08	0.25	151	3
28	0.16	0.24	152	2
29	0.23	0.24	152	2
30	0.30	0.23	153	2
31	0.37	0.23	154	2
32	0.44	0.23	154	2
33	0.51	0.22	155	2
34	0.58	0.22	156	2
35	0.65	0.22	157	2
36	0.72	0.22	157	2
37	0.80	0.22	158	2
38	0.87	0.22	159	2
39	0.95	0.22	159	2
40	1.03	0.23	160	2
41	1.11	0.23	161	2
42	1.20	0.24	162	2
43	1.29	0.24	163	2
44	1.39	0.25	164	2
45	1.49	0.26	165	3
46	1.61	0.27	166	3
47	1.75	0.29	167	3
48	1.90	0.32	169	3
49	2.09	0.35	171	4
50	2.32	0.40	173	4
51	2.63	0.47	176	5

Appendix K – Comparability Study

The following pages contain excerpted slides from a PowerPoint presentation made by the enhanced assessment grant researchers at a Technical Issues in Large Scale Assessment (TILSA) meeting on February 4, 2008. It summarizes the research completed to date. As the research project is ongoing, the final report has not yet been prepared.

Update from the North Carolina Study of the Comparability of Paper vs. Computer Based Tests

Sue Lottridge
Matt Schulz
Alan Nicewander

585 Cannery Row, Suite 201 Monterey, CA 93940 http://www.pacificmetrics.com

PACIFIC METRICS

Comparability

- Purpose of this study is two-fold:
  - Examine the comparability of scores from a Paper-based Test (PBT) and a Computer-based Test (CBT) using a hypothesis testing framework
  - Compare results from two different methodologies

PACIFIC METRICS

Six Comparability Hypotheses

- Their content must be the comparable.
- They should yield the same average scale-score, within a linear transformation.
- They should have the same measurement precision (i.e., reliability).

PACIFIC METRICS

Six Comparability Hypotheses

- They should have the same factor structure.
- They should have the same predictive validity coefficient.
- Their intercorrelation, corrected for unreliability, should be unity.

PACIFIC METRICS

Project Overview

- Two methodologies:
  - Within groups/repeated measures
    - Students take both a PBT and a CBT
    - Counterbalanced for mode order
  - Between groups
    - Students taking CBT first compared to students taking only the PBT (~65,000 students)
    - Propensity-score matching used to find comparable sample in PBT only group (results are not available)
- Two subject areas:
  - Algebra I
  - English I
- Grade 9

PACIFIC METRICS

North Carolina Tests

- Algebra 1 and English 1 Grade 9 End-of-Course tests
- All items are four-option multiple choice (no CR items)
- Test Lengths
  - Algebra 1: 64 items
  - English 1: 56 items
- Multiple forms
  - Algebra 1: 5 forms
  - English 1: 6 forms
- Two Algebra I forms were “hybrids” of other forms
- CBT forms are on-line versions of PBT forms
- CBT scale scores obtained from PBT RS-SS tables

PACIFIC METRICS

CBT characteristics

- Items presented singly
- Some scrolling required for accompanying stimulus material
- Test-taker can:
  - increase/decrease font size
  - navigate around test at will
  - click on answer choice or type in letter

PACIFIC METRICS

Test Administration

- NC administered tests in Spring 2007
- Within-subjects design
  - Students took a PBT and a CBT
  - Students randomly assigned test forms within each testing mode
  - Highest test score counted
  - Maximum of two weeks allowed between test administrations
- Volunteer schools
  - Motivators: highest test score, appeals by NC, preparation for CBT, longer testing window
  - Reasons for not participating: Lack of comfort with CBT, additional testing time
  - NC made strong efforts to obtain sample representative of state
  - Most schools were assigned a test order (PBT first, or CBT first)
  - Schools with limited computer access allowed to divide students into testing groups

PACIFIC METRICS

Data Editing

- Examinee characteristics
  - 8th and first-time 9th grade students who
  - attempted at least one item,
  - Had no administration issues,
  - Had a valid CBT administration date
  - And attended non-alternative school
- Test Order Adherence
- Form Assignment

PACIFIC METRICS

Test Order

- Test order is important:
  - Randomness in test order assignment ensures that test order effects are controlled
  - We want to compare results of CBT first group to matched sample of students who took PBT only
- Surveyed test coordinators and examined data to determine actual test order
- Data indicate that some schools did not adhere to testing order, and some schools were not assigned a testing order
  - Why might schools have chosen to test differently than assigned?

PACIFIC METRICS

Analysis of test order

- Hypothesis:
  - Test order was controlled if the previous year's test scores are the same across test order (PBT first, CBT first) within adherence levels (adhering, non-adhering, not assigned)

PACIFIC METRICS

8th Grade Math Scale Scores of Examinees, by Testing Order and School Adherence to Assigned Order

Testing Order	N	Mean	SD	*df*	*t*-Test	*d*
Schools adhering to assigned test order						
PBT First	1150	363.69	9.05	1996	-1.08	-.05
CBT First	848	363.22	9.93			
Schools NOT adhering to assigned test order						
PBT First	178	368.28	7.24	362	-10.63*	-1.12
CBT First	186	359.73	8.06			
Schools not assigned a test order						
PBT First	17	349.29	7.66	151	5.03*	1.30
CBT First	136	358.29	6.87			

* $p < .05$

PACIFIC METRICS

Data Editing Based Upon Test Order

- Algebra 1: Hypothesis was true for adhering schools, but not for other two types
  - Retain only adhering schools for analysis
- English 1: Hypothesis was true for adhering schools and for non-assigned schools, but not for non-adherers
  - Retain adhering schools and non-assigned schools for analysis

PACIFIC METRICS

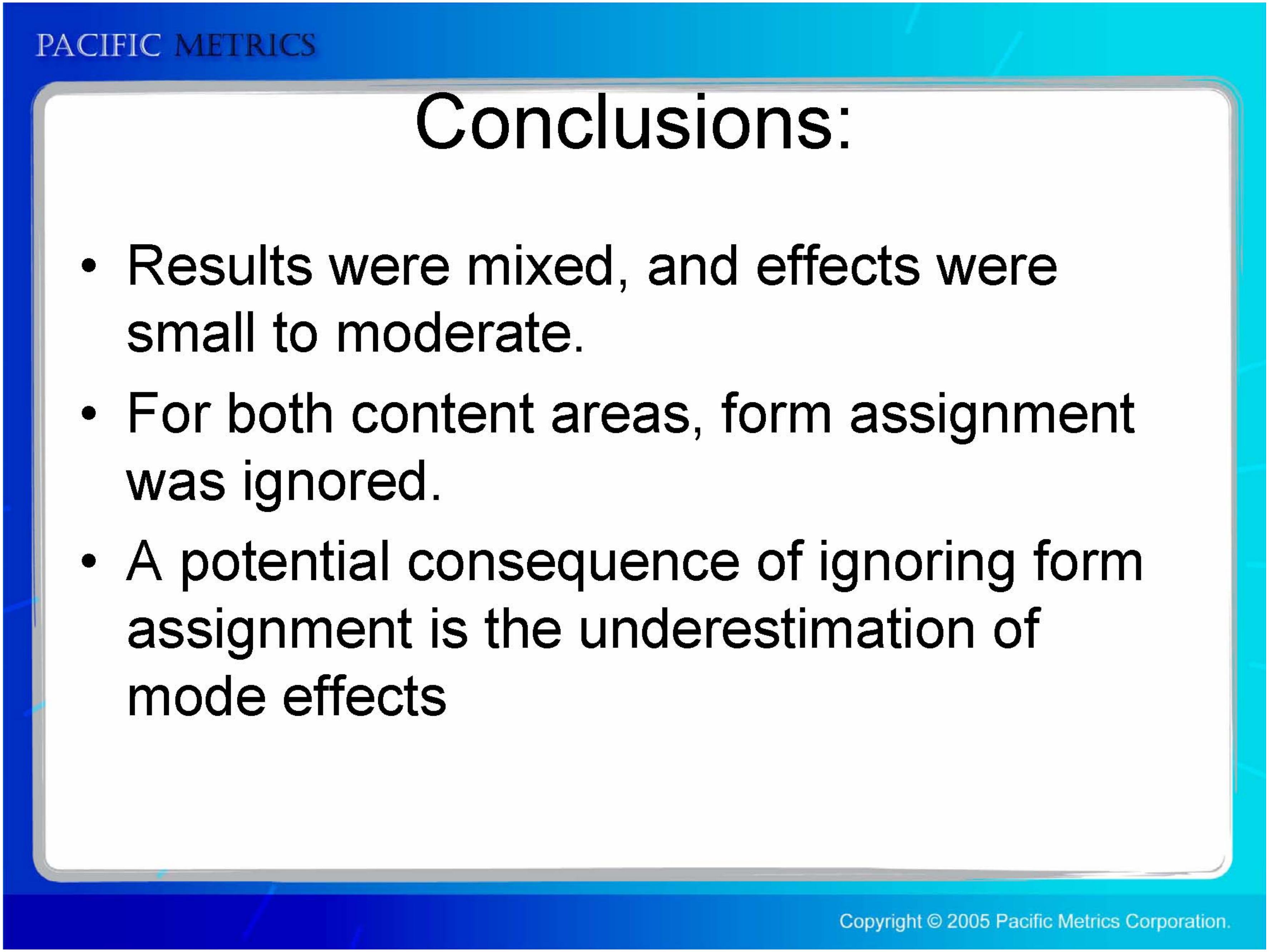
Analysis of Form Assignment

- Issues:
  - Forms were assigned randomly
  - Some students received same form in both modes
  - Some students received forms that shared some, but not all items
- Hypothesis: Scores on forms across modes should not vary by form assignment.

PACIFIC METRICS

PBT and CBT Algebra I Scale Scores by Testing Order, Mode, and Form Status

		PBT First				
		PBT		CBT		
Form Status	N	Mean	SD	Mean	SD	*d*
No items shared	336	154.36	9.92	153.66	10.26	-.07
Some items shared	610	153.89	10.03	153.82	10.36	-.01
All items shared	256	154.02	10.05	155.13	10.59	.11
		CBT First				
		PBT		CBT		
Form Status	N	Mean	SD	Mean	SD	*d*
No items shared	252	155.17	10.28	152.74	9.96	.24
Some items shared	470	155.44	11.08	152.81	11.19	.24
All items shared	177	155.54	11.43	153.12	10.98	.21


PACIFIC METRICS
Conclusions:
• Results were mixed, and effects were small to moderate.
• For both content areas, form assignment was ignored.
• A potential consequence of ignoring form assignment is the underestimation of mode effects
Copyright © 2005 Pacific Metrics Corporation.

PACIFIC METRICS

Sample sizes for analyses

	Within-subjects	PBT Only	CBT First
Algebra I	2,101	65,018	899
English I	1,527	65,648	825

PACIFIC METRICS

Analyses Conducted

- Number correct (raw) score analyses, across mode and within forms
  - Mean and standard deviations
  - Correlations
  - Reliabilities
  - Differential item functioning
- Scale score analyses, across mode
  - Means, correlations, distributions
  - Confirmatory factor analysis
  - Correlations with criterion variables
  - Assignment into achievement levels
  - Comparisons of mean scores by subgroup

PACIFIC METRICS

Comparison with PBT-Only Group

- Within-subjects and PBT-only groups somewhat different on:
  - Region of state
  - Wealth rank
  - Grade level
  - Exceptionality status
  - School type
- Groups similar on:
  - Gender and ethnicity
  - Free lunch status and LEP
  - Testing characteristics (make-up testing and accommodations)

PACIFIC METRICS

Mean raw scores and reliability

Algebra Mean Raw Scores and Reliabilities, by Mode and Form

	PBT				CBT				Differences		
Form	N	Mean	SD	α	N	Mean	SD	α	*df*	*t*-Test	*d*
F	400	42.88	12.14	.93	411	40.95	12.52	.93	809	-2.23*	-0.16
G	443	41.66	11.88	.93	420	42.20	11.86	.93	861	0.67	0.05
H	422	43.40	11.92	.93	444	42.00	12.45	.93	864	-1.69	-0.12
I	409	43.08	10.89	.91	424	41.08	11.82	.93	831	-2.54*	-0.18
J	423	42.82	11.68	.93	402	41.37	11.40	.92	823	-1.81	-0.12

PACIFIC METRICS

Raw Score Correlations

Algebra Corrected and Uncorrected Correlations by Form

		Correlations	
Form	N	*r*	r_{corr}
F	80	.92	.99
G	94	.89	.96
H	96	.93	1.00
I	75	.88	.96
J	88	.92	.99

PACIFIC METRICS

Differential Item Functioning

Classification of Items into ETS DIF Levels, by Form (64 items per form)

	No DIF		Moderate DIF		"Large" DIF	
Form	Favors PBT	Favors CBT	Favors PBT	Favors CBT	Favors PBT	Favors CBT
F	29	33	1	1	0	0
G	31	30	2	1	0	0
H	37	27	0	0	0	0
I	28	34	2	0	0	0
J	31	33	0	0	0	0
Total	156	157	5	2	0	0

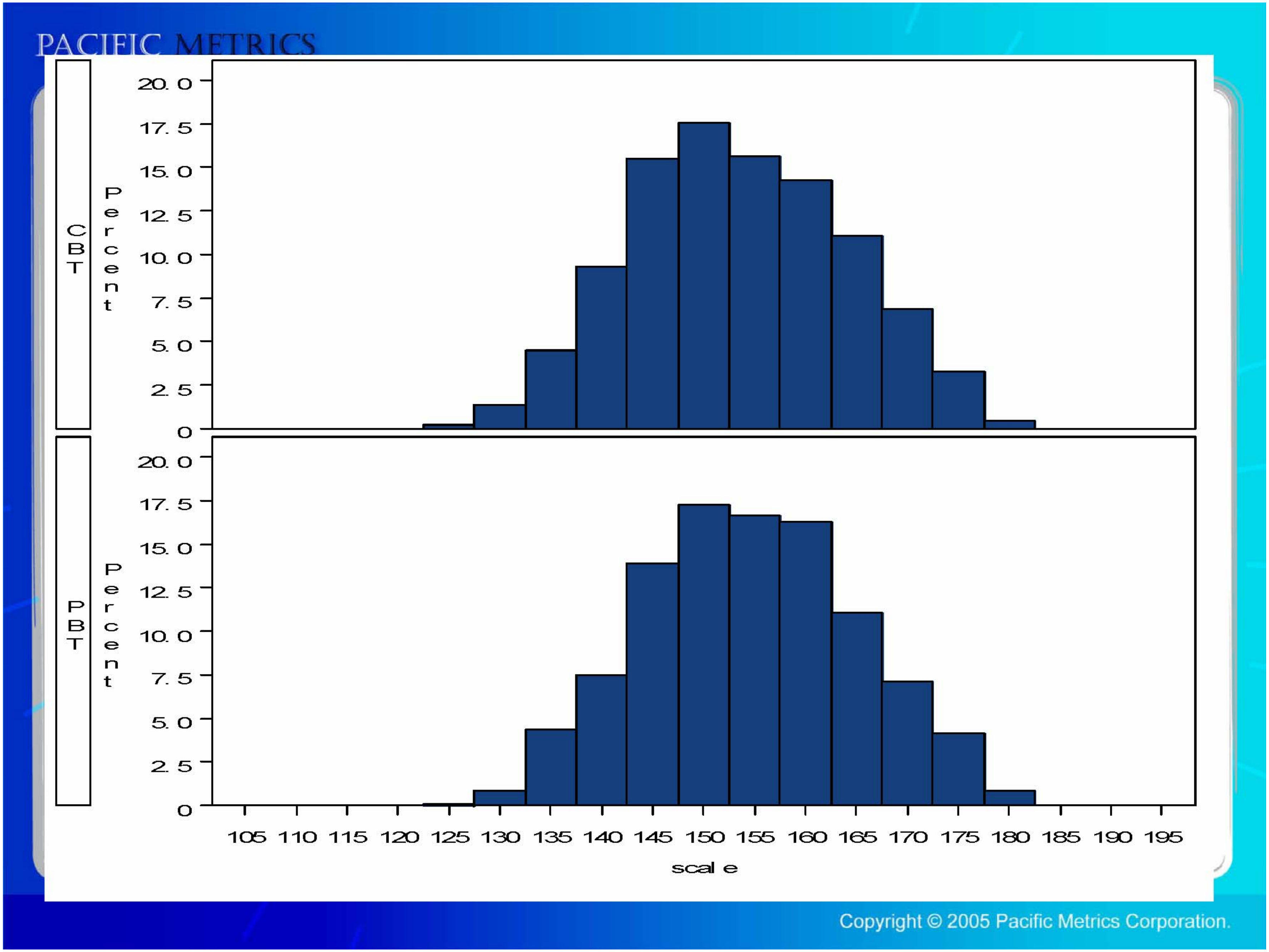
PACIFIC METRICS

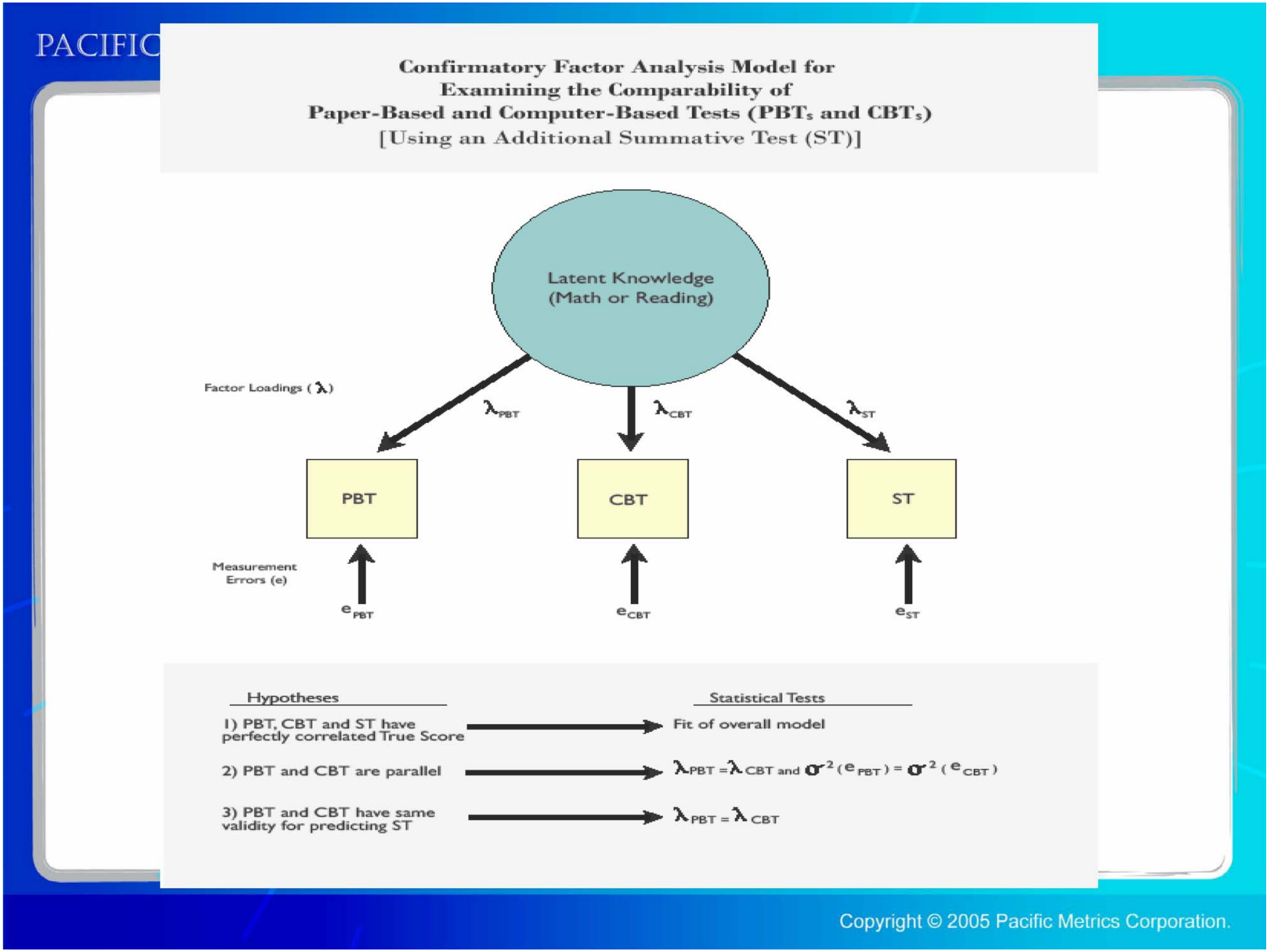
Scale score means

Algebra PBT and CBT Scale Scores by Mode (N=2,101)

Mode	Mean	SD	*df*	*t*-Test	*d*
PBT	154.62	10.42	2100	-10.43*	-.10
CBT	153.54	10.59			

*$p < .05$


PACIFIC METRICS
CBT
PBT
Percent
20.0
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0
105 110 115 120 125 130 135 140 145 150 155 160 165 170 175 180 185 190 195
scale
Copyright © 2005 Pacific Metrics Corporation.


PACIFIC
Confirmatory Factor Analysis Model for
Examining the Comparability of
Paper-Based and Computer-Based Tests (PBTs and CBTs)
[Using an Additional Summative Test (ST)]
Latent Knowledge
(Math or Reading)
Factor Loadings (λ)
λPBT
λCBT
λST
PBT
CBT
ST
Measurement
Errors (e)
ePBT
eCBT
eST
Hypotheses
Statistical Tests
1) PBT, CBT and ST have perfectly correlated True Score
Fit of overall model
2) PBT and CBT are parallel
λPBT = λCBT and σ2 (ePBT) = σ2 (eCBT)
3) PBT and CBT have same validity for predicting ST
λPBT = λCBT
Copyright © 2005 Pacific Metrics Corporation.

PACIFIC METRICS

Confirmatory Factor Analysis

Correlation-Covariance Matrix of PBT, CBT, and 8^{th} Grade Math Test Scores

	PBT	CBT	8^{th} Grade Math
PBT	*108.37*	.90	.84
CBT	99.22	*112.15*	.85
8^{th} Grade Math	82.63	85.06	*89.30*

Chi-Square and Fit Indices For Models Tested

Model	Chi Square	df	P-value	St. RMSR
Parallel Tests	6.57	2	.038	.012
Tau-Equivalent Tests	6.03	1	.014	.014

PACIFIC METRICS

Correlations with Criterion Variables

Correlations of PBT, CBT, and Criterion Variables

	7th Grade Computer Skills Test	8th Grade Math Test	8th Grade Reading Test
PBT	.68 1,889	.84 1,998	.68 1,992
CBT	.69 1,889	.85 1,998	.70 1,992

PACIFIC METRICS

Assignment into Achievement Categories for Each Mode

PBT Achievement Levels	CBT Achievement Levels I	II	III	IV	Total
I: Insufficient Mastery	106	61	4	0	171
II: Inconsistent mastery	76	241	72	0	389
III: Consistent Mastery	10	151	469	82	712
IV: Superior Performance	1	3	151	674	829
Total	193	456	696	756	2,101

Agreement	CBT	PBT (Form H)
Same Level	70.9%	71.6%
Adjacent – Higher	10.2%	14.0%
Adjacent – Lower	18.0%	14.0%
Non-Adjacent – Higher	0.2%	0.1%
Non-Adjacent – Lower	0.7%	0.1%

PACIFIC METRICS

Agreement with Teacher Rating

Assignment into Levels by PBT and Teacher Rating	PBT	CBT
Mode score same as teacher rating	56.6%	56.5%
Mode score one level higher than teacher rating	23.2%	18.2%
Mode score one level lower than teacher rating	16.5%	20.3%
Mode score two+ levels higher than teacher rating	1.6%	1.5%
Mode score two+ levels lower than teacher rating	2.2%	2.6%

PACIFIC METRICS

Sub-group differences

- No mode x sub-group interaction for:
  - Gender, ethnicity, grade-level, LEP
- Small mode x sub-group interaction for:
  - Exceptionality
    - favoring academically gifted students
  - Free Lunch Status
    - Favoring pay and temporary free lunch

PACIFIC METRICS

Algebra I Conclusions

- The computer-based tests and paper-based tests are comparable.
  - CFA indicates the tests are parallel
  - Correlations very high
  - Correlations with external variables similar
  - Mean score differences were small
  - Almost no DIF was present
  - Assignment into achievement levels was similar
- However, the CBT was slightly more difficult, and the impact of this difference should be examined:
  - What is causing this difference? Familiarity issue?
  - May need to “equate” out difference so that scores are truly interchangeable, although achievement level proportions are very similar

PACIFIC METRICS

Limitations

- Combining examinees taking same test forms/items may have biased results
- Removing examinees from non-adhering or non-assigned schools may have resulted in skewed sample
- Small sample sizes have limited ability to do some analyses:
  - Item-level dimensionality test
  - IRT calibration and equating

PACIFIC METRICS

Next Steps

- Conduct analyses on matched data

Appendix L – Achievement Level Descriptors

The following pages contain the Achievement Level Descriptors that are included in Board Policies HSP-C-010 and HSP-C-018, both of which were last updated December 6, 2007.

Achievement Level Descriptors for North Carolina End-of-Grade Tests —Pre-Grade 3 Mathematics

Level I (Pre-Grade 3)

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

Students performing at Level I show minimal understanding and computational accuracy. The students often respond with inappropriate answers or procedures. They rarely use problem-solving strategies.

Upon entering third grade, students performing at a Level I are seldom able to read, write, estimate, model, and compute using whole numbers through 999. They rarely represent and compare fractions. Students seldom recognize and use metric and customary measurement (e.g., length, temperature, time). Students infrequently identify symmetrical and congruent figures. Students lack understanding of data using Venn diagrams and pictographs. They are seldom successful when describing the results and making predictions from simple probability experiments. Level I students rarely identify and describe patterns. They seldom write accurate addition and subtraction number sentences with symbols representing unknown quantities.

Level II (Pre-Grade 3)

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

Students performing at Level II typically show some evidence of understanding and computational accuracy. The students sometimes respond with appropriate answers or procedures. They demonstrate limited use of problem-solving strategies.

Upon entering third grade, students performing at a Level II show limited proficiency when they read, write, estimate, model, and compute using whole numbers through 999. They inconsistently represent and compare fractions. Students occasionally recognize and use metric and customary measurement (e.g., length, temperature, time). At times, students identify symmetrical and congruent figures. They occasionally show understanding of data using Venn diagrams and pictographs. They conduct simple probability experiments, describe the results and make predictions with some accuracy. Level II students sometimes identify and describe patterns. They write, with limited success, addition and subtraction number sentences with symbols representing unknown quantities.

Level III (Pre-Grade 3)

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

Students performing at Level III generally show understanding, compute accurately. The students respond with appropriate answers or procedures. They use a variety of problem-solving strategies.

Upon entering third grade, students performing at Level III frequently read, write, estimate, model, and compute correctly with whole numbers through 999. They generally represent and compare fractions correctly. Students often recognize and use metric and customary measurement (e.g., length, temperature, time). Students regularly identify symmetrical and congruent figures. They show an understanding of data using Venn diagrams and pictographs. They conduct simple probability experiments, describe the results and make predictions. Level III students identify and describe patterns. They write addition and subtraction number sentences with symbols representing unknown quantities.

Level IV (Pre-Grade 3)

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

Students performing at Level IV commonly show a high level of understanding, compute accurately, and respond consistently with appropriate answers or procedures. They demonstrate flexibility by using a variety of problem-solving strategies.

Upon entering third grade, students performing at Level IV demonstrate flexibility as they read, write, estimate, model, and compute using whole numbers through 999. They accurately represent and compare fractions and also combine fractions to describe parts of a whole. Students recognize, use and apply metric and customary measurement (e.g., length, temperature, time). Students consistently identify symmetrical and congruent figures. They show an advanced understanding of data using Venn diagrams and pictographs. They conduct simple probability experiments, accurately describe the results and make predictions. Level IV students identify, describe and extend patterns. They write and apply addition and subtraction number sentences with symbols representing unknown quantities.

Achievement Level Descriptors for North Carolina End-of-Grade Tests —Grade 3 Mathematics

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

Students performing at Level I show minimal understanding and computational accuracy. The students often respond with inappropriate answers or procedures. They rarely use problem-solving strategies.

Level I students demonstrate a lack of development of number sense for whole numbers through 9,999 and a lack of evidence of ability to perform multidigit addition and subtraction. They can rarely show knowledge of multiplication facts. Students inconsistently compare, order, and represent rational numbers (halves, fourths, and eighths; thirds and sixths) concretely and symbolically. They rarely use appropriate vocabulary to compare, describe, and classify two- and three-dimensional shapes. Students are not able to correctly measure length, capacity, weight, time, and temperature (Fahrenheit and Celsius). They can sometimes identify and extend simple numeric or geometric patterns. Students show minimal understanding of organizing and displaying data using a variety of graphs. They are rarely able to identify points on rectangular coordinate system. Students seldom correctly use symbols to represent unknown quantities in number sentences and to solve simple equations. They rarely solve problems using a variety of strategies.

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

Students performing at Level II typically show some evidence of understanding and computational accuracy. The students sometimes respond with appropriate answers or procedures. They demonstrate limited use of problem-solving strategies.

Level II students show some evidence of number sense for whole numbers through 9,999 and some evidence of multidigit subtraction. They inconsistently apply multiplication facts in single-digit multiplication and division. Using fractions, they often incorrectly compare, order, and occasionally misrepresent (halves, fourths, thirds, sixths, and eighths). Students sometimes use appropriate vocabulary to compare, describe, and classify two- and three- dimensional shapes. They are inconsistent in measurement of length, capacity, weight, time, and temperature (Fahrenheit and Celsius). Students show limited understanding of the concept of probability. They are inconsistent when they identify and extend numeric and geometric patterns. Students are sometimes successful at organizing and displaying data using a variety of graphs. They sometimes correctly identify points on the rectangular coordinate system. Students occasionally correctly solve problems where symbols are used to represent unknown quantities in number sentences and to solve simple equations. They sometimes solve problems using a limited variety of strategies

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

Students performing at Level III generally show understanding, compute accurately. The students consistently respond with appropriate answers or procedures. They use a variety of problem-solving strategies.

Level III students demonstrate number sense for whole numbers through 9,999 and show consistent evidence of ability with multidigit subtraction. They know multiplication facts and are fluent with single-digit multiplication and division. They regularly are successful at comparing, ordering and representing rational numbers (halves, fourths, thirds, sixths, and eighths). Students consistently use appropriate vocabulary to compare, describe, and classify two- and three-dimensional shapes. They frequently measure length, capacity, weight, time, and temperature accurately (Fahrenheit and Celsius). Almost always, students identify and extend numeric or geometric patterns correctly. They correctly organize and display data using a variety of graphs. Students appropriately use the rectangular coordinate system to graph and identify points. They understand and use simple probability concepts.

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

Students performing at Level IV commonly show a high level of understanding, compute accurately. The students are very consistent responding with appropriate answers or procedures. They demonstrate flexibility by using a variety of problem-solving strategies.

Level IV students demonstrate a high level of success with regard to number sense for whole numbers through 9,999. They show mastery of multidigit subtraction and apply multiplication facts fluently with single-digit multiplication and division. They consistently correctly compare, order, and represent rational numbers (halves, fourths, third, sixths, and eighths). Students consistently use appropriate vocabulary to compare, describe, and classify two- and three-dimensional shapes. They accurately measure length, capacity, weight, time, and temperature (Fahrenheit and Celsius). Students successfully identify and extend complex numeric or geometric patterns. They successfully organize, display, and interpret data using a variety of graphs. Students use the rectangular coordinate system to graph, identify, and mentally manipulate points. They accurately apply simple probability concepts. Students correctly use symbols to represent unknown quantities in number sentences and to solve equations. They solve high-level thinking problems using a wide variety of strategies.

Students generally are able to use symbols to represent unknown quantities in number sentences and to solve simple equations successfully. They can solve problems using a variety of strategies.

Achievement Level Descriptors for North Carolina End-of-Grade Tests —Grade 4 Mathematics

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

Students performing at Level I show minimal understanding and computational accuracy. The students often respond with inappropriate answers or procedures. They rarely use problem-solving strategies.

Level I students rarely show number sense by comparing, ordering, estimating, and representing numbers from 0.01 to 99,999. Students are rarely able to multiply and divide multidigit numbers or use strategies for estimation of products and quotients in appropriate situations. Students are not able to add and subtract fractions with like denominators. They seldom solve problems involving the perimeter of plane figures and the area of rectangles. Students cannot make appropriate use of the coordinate plane to describe location and relative position of points. They seldom describe lines accurately as parallel or perpendicular. Students are rarely successful at collecting, organizing, analyzing, and displaying data using a variety of graphs. They are unable to use range, median, and mode to describe a set of data. Students can rarely design simple experiments to investigate and describe the probability of events. Students are unable to use the order of operations or the identity, commutative, associative, and distributive properties.

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

Students performing at Level II typically show some evidence of understanding and computational accuracy. The students sometimes respond with appropriate answers or procedures. They demonstrate limited use of problem-solving strategies.

Level II students sometimes show number sense by comparing, ordering, estimating, and representing numbers from 0.01 to 99,999. They inconsistently multiply and divide multidigit numbers. Students sometimes use strategies including estimation of products and quotients in appropriate situations. They are inconsistent in addition and subtraction of fractions with like denominators. Students sometimes solve problems involving perimeter of plane figures and the area of rectangles. Students sometimes correctly use the coordinate plane to describe the location and relative position of points. They inconsistently describe lines correctly as parallel or perpendicular. Students have difficulty collecting, organizing, analyzing, and displaying data using a variety of graphs. They are inconsistent in their ability to use range, median, and mode to describe a set of data. Students sometimes successfully design and use simple experiments to investigate and describe the probability of events. Students inconsistently use the order of operations or the identity, commutative, associative, and distributive properties.

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

Students performing at Level III generally show understanding and computational accuracy. The students consistently respond with appropriate answers or procedures. They use a variety of problem-solving strategies.

Level III students frequently show number sense by comparing, ordering, estimating, and representing numbers from 0.01 to 99,999. They are usually consistent when multiplying and dividing multidigit numbers; they use strategies including estimation of products and quotients in appropriate situations. They also add and subtract numbers with like denominators. Students solve problems involving perimeter of plane figures and area of rectangles. Students use coordinate planes to describe the location and relative position of points. They describe lines correctly as parallel or perpendicular. Students collect, organize, analyze, and display data using a variety of graphs. They use range, median, and mode to describe a set of data. Students design and use simple experiments to investigate and describe the probability of events. Students generally can use the order of operations or the identity, commutative, associative, and distributive properties.

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

Students performing at Level IV commonly show a high level of understanding and computational accuracy. The students are very consistent responding with appropriate answers or procedures. They demonstrate flexibility by using a variety of problem-solving strategies.

Level IV students successfully show number sense by comparing, ordering, estimating, and representing numbers from 0.01 to 99,999. They display fluency with multiplication and division of multidigit numbers. Students effectively use strategies including estimation of products and quotients in appropriate situations. They exhibit mastery of addition and subtraction of fractions with like denominators and decimals through hundredths. Students consistently solve problems involving the perimeter of plane figures and area of rectangles. They show a thorough understanding and application of the coordinate plane when describing location and relative position of a point. Students consistently describe lines correctly as parallel or perpendicular. They successfully collect, organize, and display data using a variety of graphs. Students accurately use range, median, and mode to describe a set of data. They effectively design and use simple experiments to investigate and describe the probability of events. Students successfully use the order of operations or the identity, commutative, associative, and distributive properties.

Achievement Level Descriptors for North Carolina End-of-Grade Tests —Grade 5 Mathematics

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

Students performing at Level I usually show minimal understanding and computational accuracy and often respond with inappropriate answers or procedures. They rarely use problem-solving strategies.

Students rarely demonstrate number sense for rational numbers 0.001 through 999,999. They rarely demonstrate ability in the addition, subtraction, comparison, and ordering of fractions and decimals. They seldom can estimate the measure of an object in one system given the measure of that object in another system. They rarely identify, estimate, and measure the angles of plane figures and rarely identify angle relationships. Students rarely identify, define, and describe the properties of plane figures, including parallel lines, perpendicular lines, and lengths of sides and diagonals. Students are seldom able to identify, generalize, and extend numeric and geometric patterns. In solving problems, fifth-graders at Level I rarely organize, analyze, and display data using a variety of graphs. They rarely are able to use range, median, and mode to describe multiple sets of data. Students rarely use algebraic expressions to solve one-step equations and inequalities. They rarely identify, describe, and analyze situations with constant or varying rates of change.

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

Students performing at Level II typically show some evidence understanding and computational accuracy and sometimes respond with appropriate answers or procedures. They demonstrate limited use of problem-solving strategies.

Students demonstrate inconsistent number sense for rational numbers 0.001 through 999,999. They demonstrate limited ability in the addition, subtraction, comparison, and ordering of fractions and decimals. They inconsistently estimate the measure of an object in one system given the measure of that object in another system. They sometimes correctly identify, estimate, and measure the angles of plane figures and sometimes correctly identify angle relationships. Students inconsistently identify, define, and describe the properties of plane figures, including parallel lines, perpendicular lines, and lengths of sides and diagonals. Students are sometimes able to identify, generalize, and extend numeric and geometric patterns. In problem solving, fifth-graders at Level II inconsistently organize, analyze, and display data using a variety of graphs. They have inconsistent success using range, median, and mode to describe multiple sets of data. Students sometimes are able to use algebraic expressions to solve one-step equations and inequalities. They inconsistently identify, describe, and analyze situations with constant or varying rates of change.

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

Students performing at Level III generally show understanding, compute accurately, and respond with appropriate answers or procedures. They use a variety of problem-solving strategies.

Students generally demonstrate number sense for rational numbers 0.001 through 999,999. They generally demonstrate ability in the addition, subtraction, comparison, and ordering of fractions and decimals. They usually make correct estimates of the measure of an object in one system given the measure of that object in another system. Students generally identify, estimate, and measure the angles of plane figures and generally identify angle relationships. They generally identify, define, and describe the properties of plane figures, including parallel lines, perpendicular lines, and lengths of sides and diagonals. Students are usually able to identify, generalize, and extend numeric and geometric patterns. To solve problems, fifth-graders at Level III generally are able to organize, analyze, and display data using a variety of graphs. They generally use range, median, and mode to describe multiple sets of data. Students generally use algebraic expressions to solve one-step equations and inequalities. They generally identify, describe, and analyze situations with constant or varying rates of change.

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

Students performing at Level IV commonly show a high level of understanding, compute accurately, and respond consistently with appropriate answers or procedures. They demonstrate flexibility by using a variety of problem-solving strategies.

Students consistently demonstrate number sense for rational numbers 0.001 through 999,999. They consistently demonstrate ability in the addition, subtraction, comparison, and ordering of fractions, mixed numbers, and decimals. They correctly estimate the measure of an object in one system given the measure of that object in another system. Students commonly identify, estimate, and measure the angles of plane figures and commonly identify angle relationships. They consistently identify, define, and describe the properties of plane figures, including parallel lines, perpendicular lines, and lengths of sides and diagonals. Students are commonly able to identify, generalize, and extend numeric and geometric patterns. To solve problems, fifth-graders at Level IV consistently organize, analyze, and display data using a variety of graphs. They consistently use range, median, and mode to describe multiple sets of data. Students commonly use algebraic expressions to solve one-step equations and inequalities. They commonly identify, describe, and analyze situations with constant or varying rates of change.

Achievement Level Descriptors for North Carolina End-of-Grade Tests —Grade 6 Mathematics

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

Students performing at Level I lack understanding and computational accuracy. They frequently respond with inappropriate answers or procedures. They seldom use problem-solving strategies.

Level I students seldom accurately add, subtract, multiply, and divide non-negative rational numbers using order of operations. They seldom correctly compare, order, and estimate with rational numbers. They lack understanding in the use of factors, multiples, exponential and scientific notation, prime factorization and percents. Level I students seldom correctly estimate and measure weight and mass of three-dimensional figures to solve problems. They seldom estimate and measure length, perimeter, area, circumference, and angles of two-dimensional figures to solve problems. They seldom can identify and describe the intersection and transformation of geometric figures in a coordinate plane. They lack understanding of counting strategies and seldom can solve problems by determining the probability of simple, compound, dependent, and independent events. Level I students seldom can simplify algebraic expressions as well as use one- and two-step equations and inequalities to represent relationships and solve problems.

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

Students performing at Level II exhibit inconsistent performance and show limited evidence of understanding. They have difficulty applying problem-solving strategies in unfamiliar situations.

Students are not consistently able to add, subtract, multiply, and divide non-negative rational numbers using order of operations. They demonstrate limited ability in the use of factors, multiples, exponential and scientific notation, prime factorization and percents. Level II students inconsistently estimate and measure weight and mass of three-dimensional figures. They inconsistently estimate and measure length, perimeter, area, circumference, and angles of two-dimensional figures to solve problems. They inconsistently identify and describe the intersection and transformation of geometric figures in a coordinate plane. Students demonstrate limited ability with counting strategies and solve problems by determining the probability of simple, compound, dependent, and independent events. They inconsistently apply algebraic principles to simplify algebraic expressions as well as use one- and two-step equations and inequalities to represent relationships and solve problems.

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

Students performing at Level III generally show understanding, compute accurately, and respond with appropriate answers or procedures. They use a variety of problem-solving strategies.

Students generally are able to accurately add, subtract, multiply, and divide non-negative rational numbers using order of operations. They usually demonstrate ability in the use of factors, multiples, exponential and scientific notation, prime factorization and percents. Students generally estimate and measure weight and mass of three-dimensional figures to solve problems. They generally estimate and measure length, perimeter, area, circumference, and angles of two-dimensional figures to solve problems. Students generally identify and describe the intersection and transformation of geometric figures in a coordinate plane. They demonstrate general ability with counting strategies and solve problems by determining the probability of simple, compound, dependent, and independent events. They generally can simplify algebraic expressions as well as use one- and two-step equations and inequalities to represent relationships and solve problems.

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

Students performing at Level IV show a high level of understanding, compute accurately, and respond consistently with appropriate answers or procedures. They demonstrate flexibility by using a variety of problem-solving strategies.

Students consistently and accurately add, subtract, multiply, and divide non-negative rational numbers using order of operations. They demonstrate fluency in the use of factors, multiples, exponential and scientific notation, prime factorization and percents. Students consistently estimate and measure weight and mass of three-dimensional figures to solve problems. They consistently estimate and measure length, perimeter, area, circumference, and angles of two-dimensional figures to solve problems. They consistently identify and describe the intersection and transformation of geometric figures in a coordinate plane. Students demonstrate fluency with counting strategies and solve problems by determining the probability of simple, compound, dependent, and independent events. They consistently are able to simplify algebraic expressions as well as use one- and two-step equations and inequalities to represent relationships and solve problems.

Achievement Level Descriptors for North Carolina End-of-Grade Tests —Grade 7 Mathematics

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

Students performing at Level I lack understanding and computational accuracy. They frequently respond with inappropriate answers or procedures. They seldom use problem-solving strategies.

Level I students show insufficient mastery of addition, subtraction, multiplication, and division of rational numbers following the order of operations. (Rational numbers may be positive, negative, or zero and include whole numbers, fractions, mixed numbers, and decimals.) Students show inability to set up and solve real-world percent problems. They rarely can write and solve proportions with rational numbers, including scaling and scale drawing. Students at Level I usually cannot solve problems involving the volume of rectangular prisms, triangular prisms, and cylinders. At Level I, students are not successful in creation of a box plot with understanding of measures of central tendency and the effect of outliers. They cannot write and solve functions from graphs, tables, or written descriptions in simpler problems. Students seldom are able to use linear equations or inequalities to solve authentic problems.

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

Students performing at Level II exhibit inconsistent performance and show limited evidence of understanding. They have difficulty applying problem-solving strategies in unfamiliar situations.

Level II students demonstrate inconsistent ability with addition, subtraction, multiplication, and division of rational numbers following the order of operations. (Rational numbers may be positive, negative, or zero and include whole numbers, fractions, mixed numbers, and decimals.) Students have difficulty with the set up and solution of real-world percent problems. They are inconsistent in ability to write and solve proportions with rational numbers, including scaling and scale drawing. Students at Level II can sometimes solve problems involving the volume of rectangular prisms, triangular prisms, and cylinders. At Level II, students are partially successful in creation of a box plot with understanding of measures of central tendency and the effect of outliers. They write and solve functions from graphs, tables, or written descriptions in simpler problems. Students can sometimes use linear equations or inequalities to solve authentic problems.

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

Students performing at Level III generally show understanding, compute accurately, and respond with appropriate answers or procedures. They use a variety of problem-solving strategies.

Level III students demonstrate consistent ability with addition, subtraction, multiplication, and division of rational numbers following the order of operations. (Rational numbers may be positive, negative, or zero and include whole numbers, fractions, mixed numbers, and decimals.) Students also show consistent ability to set up and solve real-world percent problems. They demonstrate consistent ability to write and solve proportions with rational numbers, including scaling and scale drawing. Students are able to solve problems involving the volume of rectangular prisms, triangular prisms, and cylinders. At Level III, students are usually successful in creation of a box plot with understanding of measures of central tendency and the effect of outliers. They write and solve functions from graphs, tables, or written descriptions with consistent success. Students use linear equations or inequalities to solve authentic problems.

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

Students performing at Level IV show a high level of understanding, compute accurately, and respond consistently with appropriate answers or procedures. They demonstrate flexibility by using a variety of problem-solving strategies.

Level IV students demonstrate fluency with addition, subtraction, multiplication, and division of rational numbers using order of operations. (Rational numbers may be positive, negative, or zero and include whole numbers, fractions, mixed numbers, and decimals.) Students show a high level of success to set up and solve real-world percent problems. Level IV students are very successful at writing and solving proportions with rational numbers, including scaling and scale drawing. They solve multistep surface area and volume problems including composite figures. Students consistently and accurately create a box plot from data, showing understanding of all central tendencies and the effect of outliers. They write and solve functions from graphs, tables, or written descriptions with a high level of success. Students very effectively use linear equations or inequalities to solve authentic problems.

Achievement Level Descriptors for North Carolina End-of-Grade Tests —Grade 8 Mathematics

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

Students performing at Level I lack understanding and computational accuracy. They frequently respond with inappropriate answers or procedures. They seldom use problem-solving strategies.

Level I students show lack of understanding of real numbers, including irrational numbers. They rarely are able to use indirect measurements or to use the Pythagorean Theorem to solve problems. Level I students are seldom successful at organizing and interpreting data, using scatterplots and approximating a line of best fit. Students at Level I demonstrate a lack of understanding of functions and are unable to convert functions between forms and interpret slope and intercepts. They can seldom use linear equations and inequalities to solve problems or translate between words, tables, and graphs.

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

Students performing at Level II exhibit inconsistent performance and show limited evidence of understanding. They have difficulty applying problem-solving strategies in unfamiliar situations.

Level II students show an inconsistent level of understanding of real numbers, including irrational numbers. They have difficulty using indirect measurements and using the Pythagorean Theorem to solve problems. Level II students show limited evidence of ability at organizing and interpreting data, using scatterplots and approximating a line of best fit. Students at Level II demonstrate a limited understanding of functions and are inconsistent in converting functions between forms and interpreting slope and intercepts. They have difficulty using linear equations and inequalities to solve problems, translating between words, tables, and graphs.

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

Students performing at Level III generally show understanding, compute accurately, and respond with appropriate answers or procedures. They use a variety of problem-solving strategies.

Level III students consistently show a proficient level of understanding of real numbers including irrational numbers. They generally are correct in use of indirect measurements. Students are usually successful at using the Pythagorean Theorem to solve problems. Level III students are often successful at organizing and interpreting data, using scatterplots and approximating a line of best fit. Students at Level III demonstrate an understanding of functions and can usually convert functions between forms and interpret slope and intercepts. They are generally successful at using linear equations and inequalities to solve problems, translating between words, tables, and graphs.

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

Students performing at Level IV show a high level of understanding, compute accurately, and respond consistently with appropriate answers or procedures. They demonstrate flexibility by using a variety of problem-solving strategies.

Level IV students consistently show a high level of understanding of real numbers, including irrational numbers. They correctly and accurately use indirect measurements. Students are consistently successful at using the Pythagorean Theorem to solve problems. Level IV students are highly successful at organizing and interpreting data, using scatterplots and approximating a line of best fit. Students at Level IV demonstrate a high-level understanding of functions and are successful converting functions between forms and interpreting slope and intercepts. They are highly successful at using linear equations and inequalities to solve problems, translating between words, tables, and graphs.

Achievement Level Descriptors—Algebra I EOC Test

Achievement Level I
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 show minimal conceptual understanding, limited computational accuracy, and often respond with inappropriate answers or procedures.

Students at this level are able to display data in a matrix; identify and use an element in a given matrix; add and subtract matrices; apply the laws of exponents to monomials; and add and subtract basic polynomials. Students are able to solve single-step equations and inequalities.

Achievement Level II
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 show inconsistency in conceptual understanding, computational accuracy, and in ability to respond with appropriate answers or procedures. They demonstrate limited use of problem-solving strategies.

Students at this level are able to perform basic matrix operations and interpretations; perform direct substitutions in functions and formulas; simplify formulas using order of operations; identify the greatest common factor of a polynomial; and multiply simple binomials. Students are able to use and solve two-step equations and inequalities.

Achievement Level III
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 generally show conceptual understanding and computational accuracy, and they respond with appropriate answers or procedures. They use a variety of problem-solving strategies.

Students at this level are able to write and solve linear equations; create linear models; apply and interpret constants and coefficients; understand the concepts of parallel, perpendicular and the equation of a line; solve systems of equations; factor and solve using polynomials; and use exponential and quadratic functions to solve problems. Students are able to model and solve multistep equations and inequalities.

Achievement Level IV
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 Achievement Level IV consistently show a high level of conceptual understanding, computational accuracy, and ability to respond with appropriate answers or procedures. They demonstrate capability by using a variety of problem-solving strategies.

Students at this level understand the translations of linear equations; successfully solve problems in the context of real-world situations; and interpret change in the slope, y-intercept, coefficients and constants. Level IV students are able to model and solve multistep equations and inequalities in the context of multiconcept application problems.

Algebra II End-of-Course Test—Achievement Level Descriptors

In Algebra II students apply algebraic concepts including relations, functions, polynomials, rational expressions, complex numbers, systems of equations and inequalities, and matrices. They collect and organize data to determine functions of best fit to analyze, interpret, and solve real world problems. Students use equations of circles and parabolas to model and solve problems. They model and solve problems by using direct, inverse, combined and joint variation.

Achievement Level I
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 show minimal conceptual understanding, limited computational accuracy, and often respond with inappropriate answers or procedures. They rarely use problem-solving strategies successfully.

Students at this level have difficulty applying most algebraic concepts. They have moderate success working one-step problems and those requiring routine calculator skills. The lack of computational and reasoning skills at this level results in an inability to analyze and interpret real-world problems.

Achievement Level II
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 show inconsistency in conceptual understanding, accurate computation, and responding with appropriate answers or procedures. They demonstrate limited use of problem-solving strategies.

Students at this level have some success applying algebraic concepts. They understand enough mathematical vocabulary and possess adequate calculator skills to enable them to solve problems involving more than one step. They can identify basic translations of functions, are able to perform matrix operations, can compute with complex numbers, and determine functions of best fit.

Achievement Level III
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 generally show conceptual understanding, compute accurately, and respond with appropriate answers or procedures. They use a variety of problem-solving strategies.

Students at this level are able to connect algebraic and graphical concepts. They understand and use mathematical vocabulary to solve multistep equations, inequalities, as well as systems, and are familiar with exponential and logarithmic functions. They are competent with the major features of a graphing calculator, enabling them to analyze and solve real-world problems.

Achievement Level IV

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 Achievement Level IV consistently show a high level of conceptual understanding, compute accurately, and respond with appropriate answers or procedures. They demonstrate capability by using a variety of problem-solving strategies.

Students at this level can apply algebraic and graphical concepts in a variety of contexts. They collect and organize data to determine functions of best fit to analyze, interpret, and solve real world problems. They are able to model and solve problems using various functions such as variations, quadratic, cubic, rational, and exponential. These students are able to identify several appropriate methods and techniques for solving problems.

Achievement Level Descriptors—Geometry EOC Tests

Achievement Level I

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 show minimal conceptual understanding, limited computational accuracy, and often respond with inappropriate answers or procedures. They rarely use problem-solving strategies successfully.

Achievement Level II

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 show inconsistency in conceptual understanding, computational accuracy, and in their ability to respond with appropriate answers or procedures. They demonstrate limited use of problem solving strategies and experience difficulty with complex problems and developing geometric models.

Achievement Level III

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 generally show conceptual understanding, computational accuracy, and respond with appropriate answers or procedures. They use a variety of problem-solving strategies. Students solve problems with a moderate level of complexity using one or more formulas or concepts.

Achievement Level IV

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 Achievement Level IV consistently show a high level of conceptual understanding and computational accuracy, as well as a strong ability to respond with appropriate answers or procedures. Students model and solve problems with a high level of complexity using multiple formulas or concepts.