The North Carolina Testing Program

Technical Report

2012–2015

North Carolina *EXTEND1* Alternate Assessment

(***NCEXTEND1***) Edition 3

English Language Arts/Reading Grades 3–8 and English II at Grade 10

Mathematics Grades 3–8 and Math I at Grade 10

Science Grades 5, 8 and Biology at Grade 10

Public Schools of North Carolina
State Board of Education | Department of Public Instruction

Prepared by:

Thakur Karkee, Ph. D.
Min Zhu, Ph. D.
Kinge Mbella, Ph.D.
Hope Lung, Section Chief, Test Development
North Carolina Department of Public Instruction

May 2016

Table of Contents

List of Tables

List of Figures

List of Appendices

Chapter 1 Background and Overview

1. 1 Background

It is the intent of the North Carolina (NC) General Assembly to challenge each student, including students, who have a severe intellectual disability in NC public schools with high expectations to learn, to achieve, and to fulfill his or her potential. In order to codify the intent, the General Assembly passed Bill *GCS 115C-174.10* with the following purposes for the 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."

With this mission as its guide, the North Carolina State Board of Education (SBE) developed a School-Based Management and Accountability Program to improve student performance in the early 1990s and has continually evolved its assessments and accountability systems to increase academic expectations so all students are prepared for success after high school.

In 1996, the accountability system, referred to as Accountability, Basics, and Local Control (ABCs), used data from the end-of-grade assessments to inform parents, educators and the public annually on the status of achievement at the school level. In the 1997–98 school year, five end-of-course tests were added to the ABCs school accountability model.

Since the 1990s, North Carolina has continually evolved its assessment system and its accountability system to increase academic expectations so students are prepared for success after high school. This was accomplished by re-evaluating the content standards on a 5-year cycle and based on these reviews, developing aligned assessments. Likewise, in keeping with continuous improvement, the ABCs model was amended to include associated alternate assessments and additional end-of-course assessments.

In 2005–06, ***NCEXTEND1*** alternate assessment was developed to assess students with the most significant cognitive disabilities on grade-level North Carolina *Standard Course of*

Study (NC*SCS*) Extended Content Standards in reading and mathematics at grades 3–8 and 10, science at grades 5, 8, and 10, and writing at grades 4, 7, and 10. The ***NCEXTEND1*** Edition 1 was field tested in the spring of 2006 and operationalized in 2006–2007. Analysis of the data and teacher feedback on the administration process and items resulted in the redesign and development of the ***NCEXTEND1*** Edition 2, which was first administered in 2007–2008 administration cycle.

The ABCs model continued until the 2012-13 school year when assessments aligned to the Common Cores State Standards in ELA/reading and mathematics (adopted by the SBE in June 2010) and the NC *Essential Standards* of science (adopted by the SBE in February 2010) were implemented, and the SBE adopted a new accountability model.

The North Carolina Extended Common Core for K–12 English Language Arts was adopted by the SBE in March 2011. The North Carolina Extended Common Core for K–12 Mathematics and the North Carolina Extended Essential Standards for K–12 Science were adopted by the SBE in February 2011. The Extended Content Standards are an extension of the general content standards.

Edition 3 of the ***NCEXTEND1*** assessments was implemented as a field test in the spring of 2012. In Edition 3, all items are aligned to the North Carolina Extended Content Standards (ECS) for ELA (i.e., grades 3–8, 9-10), math (i.e., grades 3-8, Algebra I A and B), and science (i.e., grades 5 and 8, high school biology). The "Algebra I A and B" course approved by the SBE has been renamed Math I although the standards for the course remain the same. The scoring model for the ***NCEXTEND1*** alternate assessment was redesigned based on test administration observations, data analysis, and teacher feedback. Students are provided up to two trials per item to select the correct response.

Beginning with 2012–13 school year, the NC READY accountability model went into effect. The test data are used for school accountability and for federal reporting but not included in growth calculation. The additional context for the current edition of the Alternate Assessments and the timeline for implementation are provided in *Table 1.1*.

This document details the design, the development, and the outcomes of the ***NCEXTEND1*** Edition 3, referred hereafter as ***NCEXTEND1***, and provides evidence on the

technical quality of the assessments. Evidences collected and documented in this report are meant to show test scores are reliable for reporting student achievement at the individual, school, district, and state levels. These evidences also support valid interpretations of test score uses described in this report.

Table 1.1 North Carolina Department of Public Instruction (NCDPI) Accountability and Testing Highlights for **NCEXTEND1** *Edition 3*

Year	Action
February 2010	The SBE adopted the NC *Essential Standards* for Science.
June 2010	The SBE adopted the NC *Standard Course of Study* (based on the *Common Core State Standards* for English Language Arts and Mathematics).
February 2011	The SBE adopted the North Carolina Extended Common Core for K–12 Mathematics and the North Carolina Extended Essential Standards for K–12 Science.
March 2011	The SBE adopted the North Carolina Extended Common Core for K–12 English Language Arts
Spring 2012	A Stand-alone field test was administered for the ***NCEXTEND1*** ELA/reading, math and science items based on the ECS.
2012–13	Beginning with 2012–13 school year, the NC *READY* Accountability Model was implemented.
Spring 2013	The operational administration of the ***NCEXTEND1*** alternate assessments of ELA/reading, math and science occurred.
July 2013	Standard setting was conducted for the ***NCEXTEND1*** alternate assessments of ELA/reading, math and science.
October 2013	The SBE adopts academic achievement standards and performance level descriptors for the ***NCEXTEND1*** alternate assessments of ELA/reading, math and science (revised by SBE action in March 2014).

1. 2 *NCEXTEND1* ELA, Math, and Science Alternate Assessments

This technical manual documents the procedures, analysis and uses for ***NCEXTEND1*** ELA/Reading and Math in grades 3–8; Science in grades 5 and 8; English II, Math I, and Biology in grade 10. The ***NCEXTEND1*** is North Carolina's alternate assessment based on alternate academic achievement standards (AA-AAAS). These assessments are designed for a portion of students with disabilities for whom the general assessment is not appropriate. The majority of students with disabilities can and should take the general assessment, with accommodations as necessary. The ***NCEXTEND1*** assessments are designed for the students identified as having the most significant cognitive disabilities (approximately 1% of the total student population). Participation for eligible students is determined by a student's Individualized Education Program (IEP). Students must be enrolled in the appropriate grade levels (3–8 and 10) to be eligible for the respective grade level ***NCEXTEND1*** assessments.

In terms of the chronology of the current Edition of the assessments, the stand-alone field tests were conducted in Spring 2012, and the operational field tests were developed for the Spring 2013 administration. The same forms have been used operationally in succeeding administrations. This technical report discusses the ***NCEXTEND1*** Edition 3 development, analysis, and results from the 2011–12 stand-alone field test to the 2014–15 operational administrations in ELA and Math grades 3–8 and grade 10 and Science grades 5, 8, and 10. The ***NCEXTEND1*** assessments are only administered to students in English and available in paper format only.

Each ***NCEXTEND1*** assessment is comprised of 15 performance-based multiple-choice items with 3 foils. These assessments are administered one-on-one to each student by an Assessor who reads aloud items and records students' responses as specified in the Assessment Guide. Students are provided up to two opportunities ("trials") to respond to each item. A student receives a score of 2 if he/she answered the item correctly during his/her first attempt. If the student selects a wrong response during his/her first trial, the Assessor following the Assessor script removes the response foil specified by the script, and the student is given a second chance to select the correct response from the two remaining foils. The student then earns a score of 1 if he/she answered the item correctly in the second trial, and 0 otherwise. *Table 1.2* shows the complete summary of total operational items and maximum possible observable score.

Table 1.2 ***NCEXTEND1****—Number of Operational Items and Maximum Possible Score Points*

Contents/Grades	Partial Credit Item		
	Number of Items	Maximum Score Points per Item	Maximum Total Score Points
ELA/Reading grades 3–8 and 10	15	2	30
Mathematics grades 3–8 and 10	15	2	30
Science grades 5, 8, and 10	15	2	30

North Carolina General Statute § 115C-174.12 mandates a statewide test administration window. The testing window for the ***NCEXTEND1*** alternate assessment is the final ten (10) instructional days of the school year. Local education agencies (LEAs)/charter schools can apply for waivers for the 2014–15 school year only and if approved, five (5) additional days was added to the test window. Exceptions are permitted to accommodate a student's Individualized Education Program (IEP) and Section 504 Plans.

1. 3 Report Summary

Chapter 1 provides a brief history of the ***NCEXTEND1*** testing in North Carolina. The chapter also describes the main features of the ***NCEXTEND1*** alternate assessments highlighting description of ***NCEXTEND1*** assessment, intended population, and administration window.

Chapter 2 presents an overview of the validation framework embedded throughout the design and development of the ***NCEXTEND1*** assessments. Validity is a unifying and core concept in test development, and thus the gathering of evidence in support of proposed uses is fundamental and should be clearly document. The first section provides a brief introduction of validity and an outline of key validity evidences as documented in this report. The second section discusses the main proposed uses of scores from the ***NCEXTEND1*** assessments.

Chapter 3 describes the 22-step test development outline adopted by the NCDPI. Key steps described in this chapter include content standards, content specification and blueprints, item development, item writer training, item review, and field test form assembly.

Chapter 4 describes the stand-alone field test administration, including the sampling plan enacted to ensure that each form was administered to a representative sample of students. In

addition, this chapter describes the audit conducted during the stand-alone field test and the steps taken to construct the operational field-test forms.

Chapter 5 of the technical report documents the procedures put in place by the NCDPI to assure the administration of the ***NCEXTEND1*** assessments are uniform, fair, and secured for all students across the state. The chapter also describes the accommodation procedures implemented to ensure every student's IEP is taken into consideration.

Chapter 6 describes the processes used for recording and scoring performance based multiple choice items and procedure adopted to ensure the quality of student data.

Chapter 7 describes the data analyses after the operational field-test administration of ***NCEXTEND1*** in 2012–13. This chapter summarizes classical test theory (CTT)-based item analysis results from the administration in 2012–13, including P-value, polyserial correlations, and Cronbach's alpha.

Chapter 8 presents a summary of the standard setting study that was conducted in July 2013 to recommend new achievement level cut scores for ***NCEXTEND1***. This chapter is a condensed version of the final report prepared by Alpine Testing Solutions, Inc. describing the full workshop and final cuts score recommendations (see Appendix 11-A).

Chapter 9 presents summary student performance results for ***NCEXTEND1*** assessments from the 2012 through 2015 administration cycles. This chapter is organized into two main sections. Section 1 highlights descriptive summary results of raw scores and reported achievement levels for ***NCEXTEND1*** across major demographic variables. Section 2 presents samples of the various standardized reports created by the NCDPI and available to local education agencies (LEAs) to provide and interpret assessments results to various stakeholders.

Chapter 10 presents summary validity evidence collected in support of the interpretation of ***NCEXTEND1*** test scores. The first couple of sections in this chapter present validity evidence in support of internal structure of these assessments. Evidence presented in these sections includes reliability, standard error of measurement estimates, and classification consistency summary of reported achievement levels. The final sections of the chapter document validity evidences: evidence based on content summarized from the alignment study, evidence based on relation to other variables, and a summary of procedures used to ensure ***NCEXTEND1*** alternate assessments are accessible and fair to all students.

Chapter 2 Validity Framework and Uses

This chapter presents an overview of the validation framework embedded throughout the design and development of the ***NCEXTEND1*** assessments. Validity is a unifying and core concept in test development, and thus the gathering of evidence in support of proposed uses is fundamental and should be clearly documented. The first section of this chapter provides a brief introduction of validity and an outline of key validity evidences. The second section discusses the proposed uses of scores from ***NCEXTEND1*** assessments.

2. 1 Summary Validation Framework for the *NCEXTEND1* Assessments

A fundamental purpose of this technical report is to present and document validity evidences on the proposed inferences of ***NCEXTEND1*** test scores as highlighted in *The Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, and National Council of Measurement in Education or AERA, APA, & NCME, 2014) hereafter referred to as the *Standards*.

> *"Validity refers to the degree to which evidences and theory support the interpretations of test scores for proposed uses of tests. Validity is, therefore, the most fundamental consideration in developing tests and evaluating tests…It is the interpretations of test scores for proposed uses that are evaluated, not the test itself."*

Standard 1.0 of the *Standards* states that *"Clear articulation of each intended test score interpretation for the specified use should be set forth, and appropriate validity evidence in support of each intended interpretation should be presented"* (p.23). Throughout this technical report, the NCDPI will be constructing, evaluating, and documenting relevant validity evidence for the proposed uses of ***NCEXTEND1*** test scores. From the test developer perspective, validation is a fluid process of evidence gathering that begins with the declaration of the proposed test use and continues throughout the life cycle of the test.

As a test developer of the ***NCEXTEND1*** alternate assessments, the NCDPI has adopted a validation framework consistent with that prescribed in the *Standards*. Under this framework, the

NCDPI is committed to ongoing evaluation of the quality of its assessments and relevance of their intended uses by continuously collecting and updating validity evidences as new data becomes available. Linn (2002, p. 46) noted that serious planning and a great deal of effort is required to accumulate evidences needed to validate the intended uses and interpretations of state assessments. His recommendation is to prioritize so that the most critical validity questions can be addressed first. *"…(W)hat are the arguments for and against the intended aims of the test? And what does the test do in the system other than what it claims?…For such questions, it is helpful to consider the level of stakes that are involved in the use or interpretation of results and then give the higher priority to those areas with highest stakes."* (Linn, 2002).

Throughout this document, validity arguments and evidences have been summarized based on prioritization of components relevant to establish the technical quality of the ***NCEXTEND1*** assessments. Even though each chapter highlights arguments and components related to particular source[s] of validity evidence, it is worth mentioning that the validation framework adapted by the NCDPI and endorsed by the *Standards* is a coherent process. A sound validity argument of the degree to which existing theory and evidence supports intended score interpretations is accomplished only by applying a holistic approach. *Table 2.1* presents an outline of the validation framework with relevant components as documented in this report.

Table 2.1 NCDPI Validation Framework for the **NCEXTEND1** *Alternate Assessments*

Sources of Validity Evidence	**References**	**Data**
Evidence based on Intended uses	Chapters 2, 9	Score report samples
Evidence based on content	Chapter 10	SEC Alignment Study Part 1
Evidence of careful test construction	Chapters 3, 4	Test construction steps, item review steps, Audit Report
Evidence based on appropriate test administration	Chapter 5	Assessment Guides
Evidence based on internal structure and reliability	Chapter 10	Cronbach alpha and SEM, Classification Consistency,
Evidence based on appropriate scoring and standard setting	Chapters 6, 7	Scoring methods, Standard Setting Report
Evidence based on careful attention to fairness for all test takers	Chapters 3, 5, 9,10	Assessment Guides
Evidence based on appropriate reporting	Chapter 9	Individual student reports, Frequency reports
Evidence based on relationship to external variables	Chapter 10	Correlation of test scores with scores from other subjects.

2. 2 Uses of *NCEXTEND1* Assessments

The North Carolina State Testing Program (NCSTP) designs, develops, and administers customized quality ***NCEXTEND1*** **alternate** assessments in grades 3–8 and grade 10. The assessments for science are aligned to NC Extended Essential Standards and those for math are aligned to the NC Extended Common Core adopted by SBE in February 2011 and those for ELA/reading aligned to NC Extended Common Core in March 2011. These assessments provide valid and reliable information intended to serve these general purposes:

- Measure ***NCEXTEND1*** students' achievement and progress to readiness as defined by the Extended Content Standards. ***NCEXTEND1*** scores are grouped and reported into 1 of 5 achievement levels (in 2012–13 scores were reported using 4 achievement levels) corresponding to 1 of the 5 performance level

descriptors adopted by the state to classify students based on their progress and readiness.

- Assessment results are used for school and district accountability under the READY Accountability Model and for federal reporting purposes. Scores from ***NCEXTEND1*** are part of the quantitative performance indicator used in two main components of the READY accountability model: annual measureable objectives (AMO) status and performance reporting. Achievement Level 3 is considered grade-level proficiency and Achievement Levels 4 and 5 are considered on track to be college and career ready.
    - AMO status indicates whether the students in the school as a whole and in each identified subgroup met the performance targets set by the state with the goal of reducing the percentage of non-proficient students. Achievement levels 4 and 5 are the proficiency standard used for AMO purposes. The number of students that an LEA deems college and career ready proficient (Levels 4 and 5) based upon alternate academic achievement standards (***NCEXTEND1***) shall not exceed 1 percent on these assessments.
    - Performance reporting identifies the percentage of students in the school who score at each of the Achievement Levels 1-5. Proficiency is reported at both grade-level proficiency (Levels 3 and above) and college and career ready proficiency (Levels 4 and 5).
- In addition to READY Accountability, North Carolina assigns School Performance Grades to schools. Assessment results from ***NCEXTEND1*** are included in School Performance Grades (SPG). The standard for proficiency in the School Performance Grades model is Achievement Level 3 and above.

Standard 1.1 of the *Standards* (AERA, APA & NCME, 2014) states that *"Test developers should set forth clearly how test scores are intended to be interpreted and consequently used …"* (p. 23). To this end, the NCDPI provides score reports, along with interpretative guidance, at the student, school, district, and state levels. The interpretative guidance helps stakeholders at the

classroom, school, and district levels understand the content and uses of these reports. These guides are also intended to help administrators and educators explain test results to parents and the general public. To ensure the ***NCEXTEND1*** assessment scores are used as intended, the North Carolina *Testing Code of Ethics* (see Appendix 2-A Testing Code of Ethics) provides guidance on the appropriate use of test scores and reports. One intent of the *Testing Code of Ethics* is to help educators recognize that a test score is only a single piece of information and must be interpreted as intended. This is at the core of validity: it is the intended interpretation[s] of test scores which are valid, not the test itself.

WinScan is a software application provided to LEAs (available to test coordinators at the district level) by the NCDPI that is used to generate a variety of score reports: class roster reports, score frequency reports, achievement level frequency reports.

Table 2.2 shows a list of reports described in Chapter 8. The individual student reports (ISRs) are designed for students and parents, teachers, and school administrators. Class rosters are designed for teachers and school administrators. Score and achievement level frequency reports are designed for teachers, school administrators, district administrators, and state administrators.

Table 2.2 WinScan **NCEXTEND1** *Reports and Intended Audience*

Report	***Audience***				
	Parent	**Administrators**			
		Teacher	**School**	**District**	**State**
Individual Student Report (ISRs)	✓	✓	✓		
Class Roster Reports		✓	✓		
Score and Achievement Level Frequency Reports		✓	✓	✓	✓

2. 3 Confidentiality of Student Test Scores

State Board of Education policy GCS-A-010 (j)(1) states, *"Educators shall maintain the confidentiality of individual students. Publicizing test scores or any written material containing personally identifiable information from the student's educational records shall not be disseminated or otherwise made available to the public by a 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 3 Field Test Design and Development Process

Standard 4.0 of the *Standards* (AERA, APA, & NCME, 2014) states *"...Test developers and publishers should document steps taken during the design and development process to provide evidence of fairness, reliability, and validity for intended uses for individuals in the intended examinee population"* (p. 85). In adherence with the *Standards*, this chapter documents steps implemented by the NCDPI during design and development of ***NCEXTEND1*** assessments. Key aspects of design and development described in this chapter include content specification and blueprints, item development, and item review. *Figure 3.1* shows the 22 steps test development flow chart prescribed by the North Carolina State Board of Education (SBE; 2003, 2012). According to SBE policy (2012):

> *...the state-adopted content standards are periodically reviewed for possible revisions; however, test development is continuous. The NCDPI Accountability Services/Test Development Section test development staff members begin developing* ***operational*** *test forms for the North Carolina Testing Program when the State Board of Education determines that such tests are needed. The need for new tests may result from mandates from the federal government or the North Carolina General Assembly. New tests can also be developed if the SBE determines the development of a new test will enhance the education of North Carolina students. The test development process consists of six phases and takes approximately four years. The phases begin with the development of test specifications and end with the reporting of operational test results.*

Additional information regarding North Carolina state assessment development process including standard setting, alignment, and test development can also be found on the NCDPI/Technical Notes web page. The test development process (Table *3.1*) Steps 3 to 7 only applies to the general tests with technology-enhanced items. Since ***NCEXTEND1*** items are performance-based multiple-choice items, the item tryout steps were not part of the ***NCEXTEND1*** test development process.

Table 3.1 Test Development Process Flow Chart

Adopt Content Standards		Step 8 Develop New Items	Step16 Review Assembled Test
Step 1[1] Develop Test Specifications (Blueprint)		Step 9[2] Review Items for Field Test	Step17 Final Review of Test
Step 2[2] Develop Test Items		Step 10 Assemble Field Test Forms	Step 18[12] Administer Test as Pilot
Step 3[1] Review Items for Tryouts		Step 11 Review Field Test Forms	Step19 Score Test
Step 4 Assemble Item Tryout Forms		Step 12[2] Administer Field Test	Step 20[12] Establish Standards
Step 5 Review Item Tryout Forms		Step 13 Review Field Test Statistics	Step 21[2] Administer Test as Fully Operational
Step 6[2] Administer Item Tryouts		Step14[2] Conduct Bias Reviews	Step 22 Report Test Results
Step 7 Review Item Tryout Statistics		Step15 Assemble Operational Test	

[1] Activities done only at implementation of new curriculum

[2] Activities involving NC teachers

3.1 Content Standards and Curriculum Connectors

As stated in Chapter 1 (see Table 1.1), the North Carolina Extended Content Standards are aligned to the *Common Core State Standards* (*CCSS*) and the North Carolina *Essential Standards*. Operational field test forms aligned to the Extended Content Standards were administered in the 2012–13 testing administration (READY initiative). To ensure items written for the NC assessments (regular and alternate) met the cognitive rigor as specified in the adopted standards, the NCDPI Test Development section worked with the NCDPI Curriculum and Instruction Division to provide training workshops on Revised Bloom Taxonomy (RBT), depth of knowledge (DOK), and overall alignment of assessments to content standards.

3.1.1 Revised Bloom Taxonomy and Depth of Knowledge

As part of pre-item development training, the NCDPI Test Development section with collaboration from the NCDPI Curriculum and Instruction Division organized two main workshops on RBT and Webb's DOK. The first workshop was organized on July 8, 2010. The focus of the workshop was to get NCDPI test measurement specialist (TMS), North Carolina State University Technical Outreach for Public Schools (NCSU-TOPS) content leads, and NCDPI Curriculum and Instruction content specialists familiarized with Hess's matrix, which the NCDPI had decided to use for alignment purposes because it relates RBT to Webb's alignment scheme. Karin Hess (researcher at Center for Assessment) developed a four-by-six table containing Webb's DOK levels across the top and RBT process dimension across the side (see *Table 3.2*). During the workshop, participants received training and started to classify NC*SCS* and NC*ESS* using Hess's matrix.

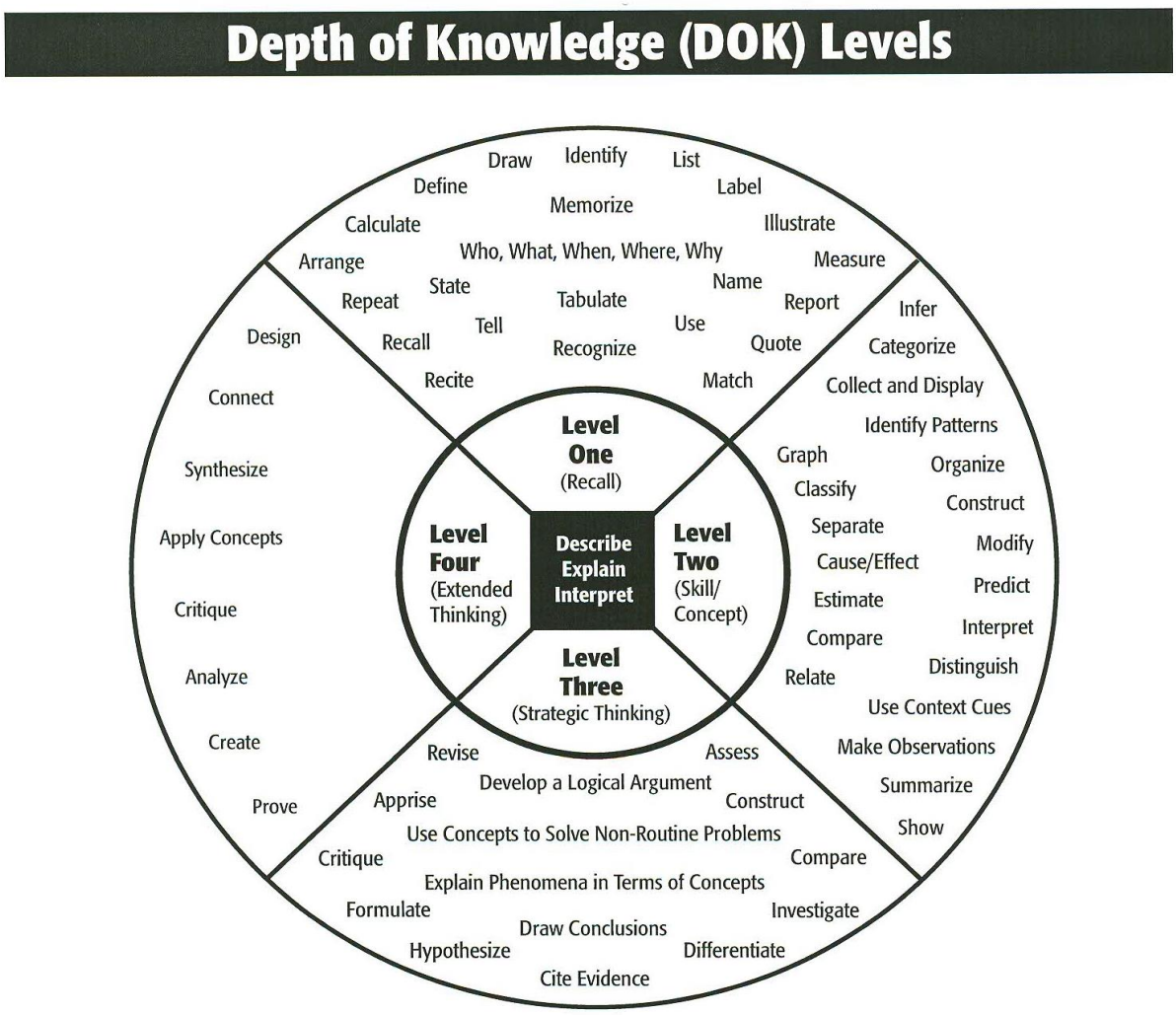
Table 3.2 Hess' Cognitive Rigor Matrix with Curricular Examples

Bloom's Revised Taxonomy of Cognitive Process Dimensions	**Webb's Depth-of-Knowledge (DOK) Levels**			
	Level 1 **Recall & Reproduction**	**Level 2** **Skills & Concepts**	**Level 3** **Strategic Thinking/Reasoning**	**Level 4** **Extended Thinking**
Remember Retrieve knowledge from long-term memory, recognize, recall, locate, identify	○ Recall, recognize, or locate basic facts, ideas, principles ○ Recall or identify conversions between representations, numbers, or units of measure ○ Identify facts/details in texts			
Understand Construct meaning, clarify, paraphrase, represent, translate, illustrate, give examples, classify, categorize, summarize, generalize, infer a logical conclusion (such as from examples given), predict, compare/contrast, match like ideas, explain, construct models	○ Compose & decompose numbers ○ Evaluate an expression ○ Locate points (grid/ number line) ○ Represent math relationships in words, pictures, or symbols ○ Write simple sentences ○ Select appropriate word for intended meaning ○ Describe/explain how or why	○ Specify and explain relationships ○ Give non-examples/examples ○ Make and record observations ○ Take notes; organize ideas/data ○ Summarize results, concepts, ideas ○ Make basic inferences or logical predictions from data or texts	○ Explain, generalize, or connect ideas using supporting evidence ○ Explain thinking when more than one response is possible ○ Explain phenomena in terms of concepts ○ Write full composition to meet specific purpose	○ Explain how concepts or ideas specifically relate to other content domains or concepts ○ Develop generalizations of the results obtained or strategies used and apply them to new problem situations
Apply Carry out or use a procedure in a given situation; carry out (apply to a familiar task), or use (apply) to an unfamiliar task	○ Follow simple/routine procedure (recipe-type directions) ○ Solve a one-step problem ○ Calculate, measure, apply a rule ○ Apply an algorithm or formula (area, perimeter, etc.) ○ Represent in words or diagrams a concept or relationship ○ Apply rules or use resources	○ Select a procedure according to task needed and perform it ○ Solve routine problem applying multiple concepts or decision points ○ Retrieve information from a table, graph, or figure and use it solve a problem requiring multiple steps ○ Use models to represent concepts ○ Write paragraph using	○ Use concepts to solve non- routine problems ○ Design investigation for a specific purpose or research question ○ Conduct a designed investigation ○ Apply concepts to solve non- routine problems ○ Use reasoning, planning, and evidence ○ Revise final draft for meaning	○ Select or devise an approach among many alternatives to solve a novel problem ○ Conduct a project that specifies a problem, identifies solution paths, solves the problem, and reports results ○ Illustrate how multiple themes (historical, geographic, social) may be interrelated
Analyze Break into constituent parts, determine how parts relate, differentiate between relevant-irrelevant, distinguish, focus, select, organize, outline, find coherence, deconstruct (e.g., for bias or point of view)	○ Retrieve information from a table or graph to answer a question ○ Identify or locate specific information contained in maps, charts, tables, graphs, or diagrams	○ Categorize, classify materials ○ Compare/contrast figures or data ○ Select appropriate display data ○ Organize or interpret (simple) data ○ Extend a pattern ○ Identify use of literary devices ○ Identify text structure of paragraph ○ Distinguish: relevant-irrelevant information,	○ Compare information within or across data sets or texts ○ Analyze and draw conclusions from more complex data ○ Generalize a pattern ○ Organize/interpret data: complex graph ○ Analyze author's craft,	○ Analyze multiple sources of evidence or multiple works by the same author, or across genres or time periods ○ Analyze complex/abstract themes ○ Gather, analyze, and organize information
Evaluate Make judgments based on criteria, check, detect inconsistencies or fallacies, judge, critique			○ Cite evidence and develop a logical argument for concepts ○ Describe, compare, and contrast solution methods ○ Verify reasonableness of results	○ Gather, analyze, & evaluate relevancy & accuracy ○ Draw & justify conclusions ○ Apply understanding in a novel way, provide argument
Create Reorganize elements into new patterns/structures, generate, hypothesize, design, plan, construct, produce	○ Brainstorm ideas, concepts, or perspectives related to a topic or concept	○ Generate conjectures or hypotheses based on observations or prior knowledge	○ Synthesize information within one source or text ○ Formulate an original problem given a situation ○ Develop a complex model for a given situation	○ Synthesize information across multiple sources or texts ○ Design a model to inform and solve a real-world, complex, or abstract situation

In July 2010, NCDPI organized a one-day face-to-face training session on Webb's alignment method. Norm Webb was invited to facilitate the training on alignment and DOK. During the first four hours of the training, Webb presented an overview of his alignment model (Webb et. al., 2005) and his definitions of Depth-of-Knowledge (see *Figure 3.1*). Slides used for the training are in Appendix 3-A Norm Webb Training–Content Complexity.

This workshop was built on the July 8, 2010, workshop in which participants were able to classify standards using the Hess matrix. During the July 26, 2010, workshop, participants received training on aligning items using the RBT framework and how to classify items based on their cognitive complexity using the Webb alignment tool, which organizes verbs into general DOK categories.

Figure 3.1 Webb Alignment Tool



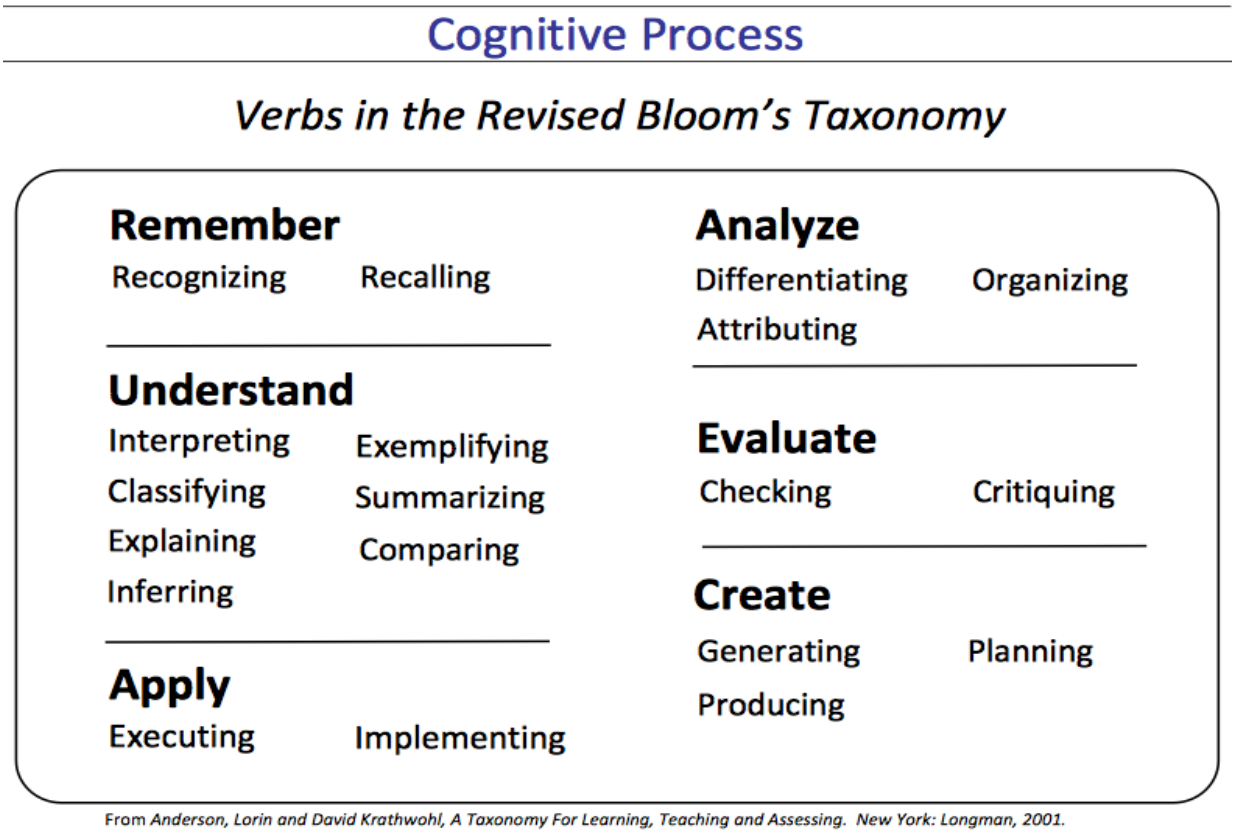
Level One Activities	Level Two Activities	Level Three Activities	Level Four Activities
Recall elements and details of story structure, such as sequence of events, character, plot and setting.	Identify and summarize the major events in a narrative.	Support ideas with details and examples.	Conduct a project that requires specifying a problem, designing and conducting an experiment, analyzing its data, and reporting results/ solutions.
Conduct basic mathematical calculations.	Use context cues to identify the meaning of unfamiliar words.	Use voice appropriate to the purpose and audience.	Apply mathematical model to illuminate a problem or situation.
Label locations on a map.	Solve routine multiple-step problems.	Identify research questions and design investigations for a scientific problem.	Analyze and synthesize information from multiple sources.
Represent in words or diagrams a scientific concept or relationship.	Describe the cause/effect of a particular event.	Develop a scientific model for a complex situation.	Describe and illustrate how common themes are found across texts from different cultures.
Perform routine procedures like measuring length or using punctuation marks correctly.	Identify patterns in events or behavior.	Determine the author's purpose and describe how it affects the interpretation of a reading selection.	Design a mathematical model to inform and solve a practical or abstract situation.
Describe the features of a place or people.	Formulate a routine problem given data and conditions.	Apply a concept in other contexts.	
	Organize, represent and interpret data.		

Webb, Norman L. and others. "Web Alignment Tool" 24 July 2005. Wisconsin Center of Educational Research. University of Wisconsin-Madison. 2 Feb. 2006. <http://www.wcer.wisc.edu/WAT/index.aspx>.

3.1.2 Curriculum Development

North Carolina uses the RBT to help educate students on the complex thinking skills expected of 21st Century graduates. The RBT was chosen because it has well-defined verbs and is based on modern cognitive research. The RBT categorizes both the **cognitive process** (*Figure 3.2*) and the **knowledge dimension** of the standards. The cognitive process is delineated by the verb used in the standards. The chart below illustrates the verbs used in the RBT and their specific definitions.

Figure 3.2 Cognitive Process: Verbs in the Revised Bloom's Taxonomy



From *Anderson, Lorin and David Krathwohl, A Taxonomy For Learning, Teaching and Assessing. New York: Longman, 2001.*

A common understanding of these verbs by teachers is the backbone of professional development around the new standards. The knowledge dimension is a way to categorize the type of knowledge to be learned. For instance, in the standard *"the student will understand the concept of equality as it applies to solving problems with unknown quantities,"* the knowledge to be learned is *"the concept of equality as it applies to solving problems with unknown quantities."* Knowledge in the RBT falls into four categories:

- Factual Knowledge
- Conceptual Knowledge
- Procedural Knowledge
- Meta-Cognitive Knowledge

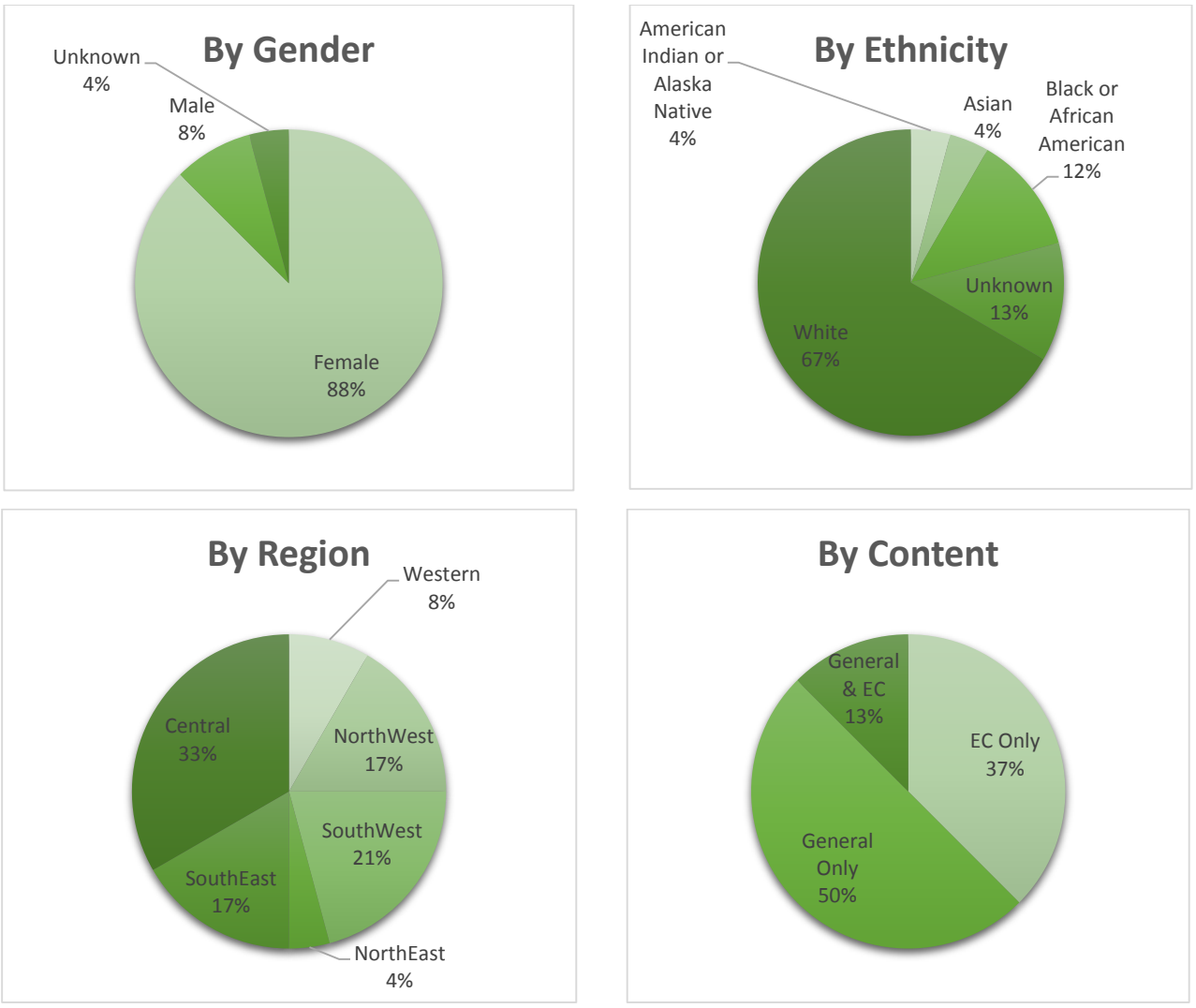
3.2 Step 1. Content Domain Specification and Blueprints

Test specifications for the ***NCEXTEND1*** assessments were developed with the focus on content specified in the Extended Content Standards. It was determined that the test blueprint would be developed using the goal percentages that align with the end-of-grade assessments for ELA/reading and mathematics at grades 3–8 and science at grades 5 and 8 and the end-of-course assessments of English II, Math I, and Biology to the maximum extent possible. Standard 4.1 of the *Standard* (AERA, APA, & NCME, 2014) states:

> *"Test specifications should describe the purpose(s) of the test, the definition of the construct or domain measured, the intended examinee population, and interpretations for intended uses. The specifications should include a rationale supporting the interpretations and uses of test results for the intended purpose(s)"* (p. 85).

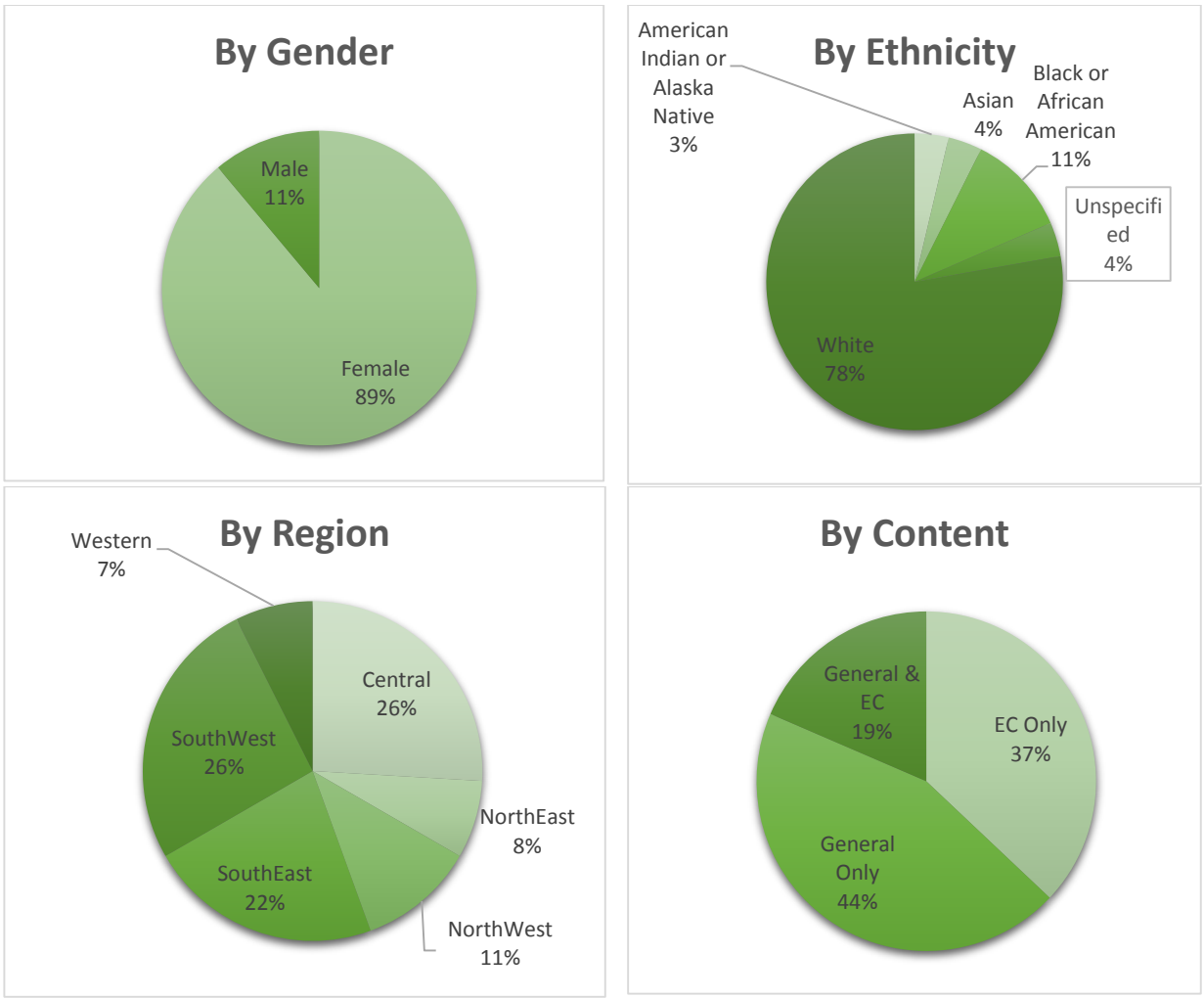
In addition, Standard 4.12 of the *Standard* (AERA, APA, & NCME, 2014) states, "*Test developers should document the extent to which the content domain of a test represents the domain defined in the test specifications*" (p. 89). The NCDPI invited teachers to collaborate and develop recommendations for a prioritization of the standards indicating the relative importance of each standard, the anticipated instructional time, and the appropriateness of the standard to performance-based multiple-choice item type. *Figures 3.3* through *3.5* present demographic information of educators who participated in the prioritization of the extended content standard for ELA, mathematics, and science respectively. In all, 24 educators were on the ELA panel, 27 educators were on the mathematics panel, and 31 educators were on the science panel. As shown in *Figures 3.3* through *3.5*, the majority of teachers were white females, and nearly half of the panel taught general education. The general education teachers in the panel strengthened the items alignment with the Extended Content Standards.

Figure 3.3 Demographic Information of Extended Content Standards Prioritization Panel—ELA



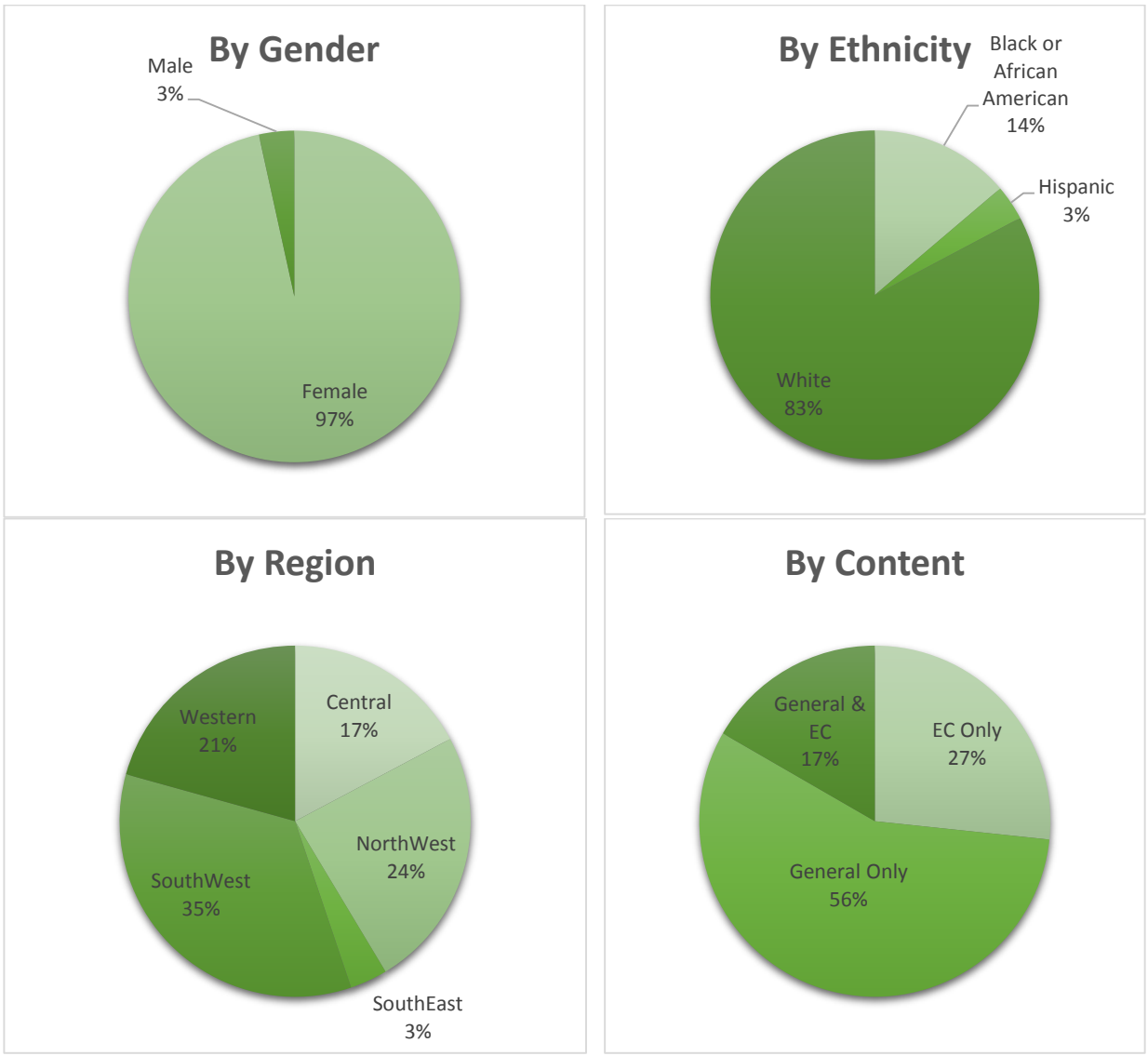
Note: N=24: General: General Education; EC: Exceptional Children.

Figure 3.4 Demographic Information of *Extended Content* Standards Prioritization Panel—Math



Note: N=27: General: General Education; EC: Exceptional Children.
Some teachers taught both general and exceptional children ELA classes.

Figure 3.5 Demographic Information of Extended Content Standards Prioritization Panel—Science



Note: N=31: General: General Education; EC: Exceptional Children.
Some teachers taught both general and exceptional children ELA classes.

Subsequently, curriculum and test development staff from the NCDPI reviewed the recommendations from the teacher panels and developed weighted distributions of the number of items sampled across domains for each grade level. Based on the content domain specification, test blueprints were developed that matched the number of items from each assessable standard to be represented on each test form. Table 3.3 through Table 3.5 show the sampling of the ELA, Math, and Science adopted content domain specifications respectively

for the ***NCEXTEND1*** alternate assessments based on the Extended Content Standards. As an example of how to read the information in these tables, there are three rows within Grade 3 ELA (see *Table 3.3*). The values in the "Blueprint" row show the suggested weights in percentage for each content standard within ELA, the row "No. of items" shows the number of items represented in the test from each content standard on the 15-item test, and the "Actual" row shows the actual weights in percentage once the final form was constructed.

Table 3.3 **NCEXTEND1** *ELA Grades 3–8 and 10 Content Standards and Weights*

Grade		Reading Literature (RL)	Reading Information (RI)	Language (L)	Total
3	Blueprint	33	40	27	100
	No. of Items	5	6	4	15
	Actual	33.3	40.0	26.7	100
4	Blueprint	33	40	27	100
	No. of Items	5	6	4	15
	Actual	33.3	40.0	26.7	100
5	Blueprint	33	40	27	100
	No. of Items	5	6	4	15
	Actual	33.3	40.0	26.7	100
6	Blueprint	33	40	27	100
	No. of Items	5	6	4	15
	Actual	33.3	40.0	26.7	100
7	Blueprint	33	40	27	100
	No. of Items	5	6	4	15
	Actual	33.3	40.0	26.7	100
8	Blueprint	33	40	27	100
	No. of Items	5	6	4	15
	Actual	33.3	40.0	26.7	100
English II	Blueprint	33	46	20	99
	No. of Items	5	7	3	15
	Actual	33.3	46.7	20.0	100

Table 3.4 shows the blueprint weights in percentage, number of items, and actual weight in percentage for math. Grades 3–5 measures the same standards with varying weights as grade level changes. The content standards for grade 6–8 emphasize "Expressions and equations" and "Statistics and probability." Grade 10 standards place emphasis on Algebra.

Table 3.4 **NCEXTEND1** *Mathematics Grades 3–8 and 10 Content Standards and Weights*

Grades 3-5		Operations and Algebraic Thinking	Numbers and Operations in Base Ten	Numbers and Operations-Fractions	Measurement and Data	Geometry	Total
3	Blueprint	33	20	7	27	13	100
	No. of Items	5	3	1	4	2	15
	Actual	33.3	20.0	6.7	26.7	13.3	100
4	Blueprint	40	7	20	20	13	100
	No. of Items	6	1	3	3	2	15
	Actual	40.0	6.7	20.0	20.0	13.3	100
5	Blueprint	20	40	13	13	13	99
	No. of Items	3	6	2	2	2	15
	Actual	20	40	13.3	13.3	13.3	100
Grades 6-8		Ratios and Proportional Relationships	The Number System	Expressions and Equations	Geometry	Statistics and Probability	Total
6	Blueprint	20	40	13.3	13.3	13.3	99.9
	No. of Items	3	6	2	2	2	15
	Actual	20	40	13.3	13.3	13.3	100
7	Blueprint	13	20	40	13	13	99
	No. of Items	2	3	6	2	2	15
	Actual	13.3	20.0	40.0	13.3	13.3	100
8	Blueprint			40	47	13	100
	No. of Items			6	7	2	15
	Actual			40	46.7	13.3	100
Grade 10		Number and Quantity: The Real Number System	Number and Quantity: Quantity	Algebra: See structure in Expression	Algebra: Creating Equations	Algebra: Reasoning with Equations and Inequalities	Total
Math I	Blueprint	47	7	13	13	20	100
	No. of Items	7	1	2	2	3	15
	Actual	46.7	6.7	13.3	13.3	20	100

For science *Table 3.5* shows the blueprint weights in percentage, number of items, and actual weight in percentage. In Biology, the focus is primarily on living organism and ecosystem.

Table 3.5 **NCEXTEND1** *Science Grades 5 and 8 and 10 Content Standards and Weights*

Grade		Forces and Motion (P1)	Matter, Properties, and Change (P2)	Earth Systems, Structures, and Processes (E1)	Structures and Functions of Living Organisms (L1)	Ecosystem (L2)	Total
5	Blueprint	13	13	20	27	27	100
	No. of Items	2	2	3	4	4	15
	Actual	13.3	13.3	20.0	26.7	26.7	100
8	Blueprint	13.0	7.0	27.0	27.0	27.0	101
	No. of Items	2	1	4	4	4	15
	Actual	13.3	6.7	26.7	26.7	26.7	100
Biology	Blueprint				33	67	100
	No. of Items				5	10	15
	Actual				33.3	66.7	100

3.3 Step 2. Item Development

In Step 2, the NCDPI began the process of writing and aligning items to NC grade-level ***NCEXTEND1*** assessments blueprints. This section as well as Sections 3.4 and 3.5 discusses item development as prescribed in Standard 4.7 of the *Standards* (AERA, APA, & NCME, 2014), which states, *"The procedures used to develop, review, and try out items and to select items from the item pool should be documented"* (p. 87). The Extended Content Standards provides every assessed content area a set of competencies for each grade. The intent is to ensure rigorous academic content standards that are uniform across the state. This effort is based on a philosophy of teaching and learning that is consistent with current research, exemplary practices, and national standards. Items were written using a plain English approach to align with specific grade-level objectives from the Extended Content Standards as defined in the test blueprint. .

3.3.1 Plain English Approach

Prior to the development of items, the NCDPI on April 28, 2011, conducted a workshop on the use of "Plain English" practices in test construction. The workshop was facilitated by Dr. Edynn Sato, director of Research and English Learner Assessment with the Assessment and Standard Development Services Program at West Ed. Participants at this work included personnel from the NCDPI Division of Accountability Services (including the test development

section), Curriculum and Instruction Division, and NCSU-TOPS staff. The one-day training workshop focused on the latest research in the area of plain English practices and examined its use in the NCDPI training for item writers and reviewers. Lessons learned from this training were used to re-evaluate how items for the new assessments were developed following the plain English framework, which emphasize clarity without altering the construct being assessed. In general, the goal was to develop items that assess the construct without adding in construct-irrelevant variance that may come into play if the students cannot access and interpret what is being required of them.

The training emphasized aspects of the test items, such as presentation of material, socio-cultural contexts, and culture-specific references, which may interfere with the measurement of the student's ability to demonstrate their knowledge of the content. This is also known as construct-irrelevant variance. Such construct-irrelevant variance can lead to an underestimation of the student's true ability level. Strategies such as Universal Design and Plain English have been found to increase access by reducing unnecessary linguistic and cultural complexities, thus reducing construct-irrelevant variance for students for which these factors may exist while still maintaining appropriate measurement of the construct for the entirety of the student population. These core principles were emphasized in the item writer training courses designed by NCDPI and required to be taken by all potential item writers/reviewers. The complete workshop materials including the workshop agenda is available in Appendix 3-B Exhibit 307 Plain English Training.

3.3.2 Development of *NCEXTEND1* Items

It was the goal of the NCDPI that NC teachers would be trained as ***NCEXTEND1*** item writers and reviewers in all grade levels. Staff from NCSU-TOPS who were integral in the development of the North Carolina Alternate Assessment Portfolio (North Carolina's first alternate assessment) and Edition 1 of the ***NCEXTEND1***, and who also have expertise regarding the special needs of this student population worked with staff from the NCDPI Division of Exceptional Children to provide item writing training to content specialists at NCSU-TOPS and teachers from across the state. Sample materials used in the training are attached in Appendix

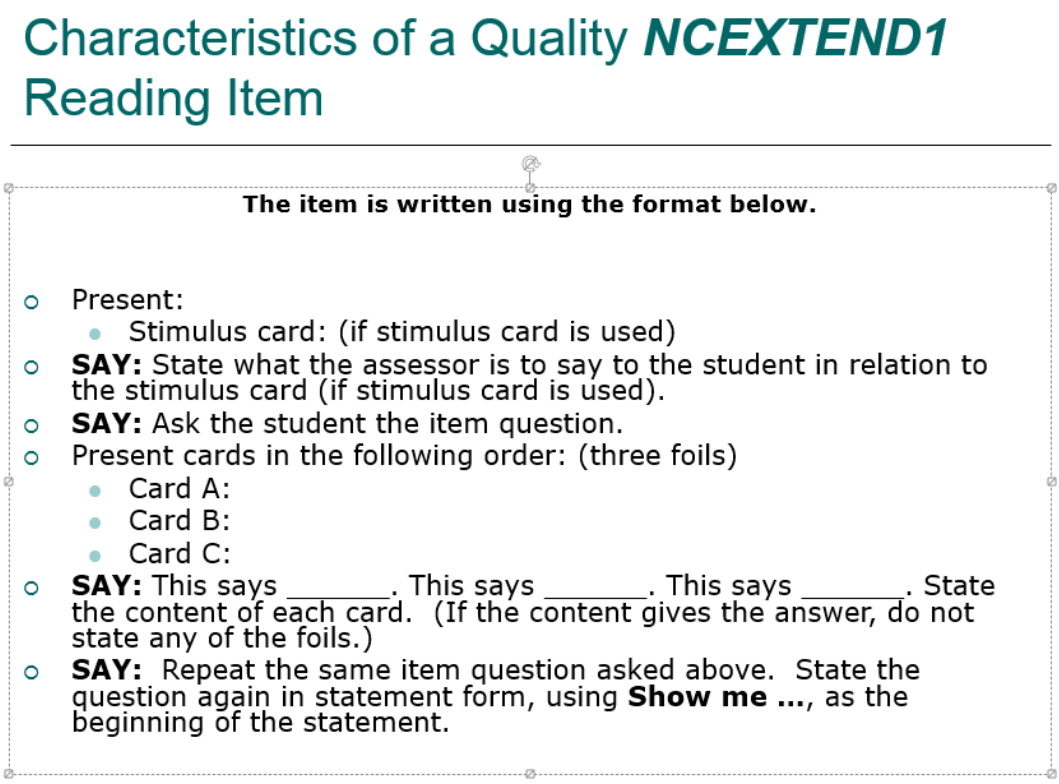
3-C. Those staff members who provided the training also participated in item writing sessions held by each NCDPI content area team.

The ***NCEXTEND1*** follows an on-demand question and answer format. Each item page contains a list of the materials to be used from the Manipulatives Kit, directions and script for item presentation, and specific scoring criteria. The materials provided in the Manipulatives Kit represent the information that students actually see during the test administration. Items are presented in the way similar to how flash cards are used. Manipulatives represent a combination of flash cards with text only, pictures and text, and pictures only. An example of the NCEXTEND1 test item is shown in Appendix 3-D.

3.3.3 Item Writer Training

The current ***NCEXTEND1*** is marked by several changes to the assessment's design, development, administration, and scoring processes. The goal was to train teachers on how to effectively write quality performance-based, multiple-choice items with a single stem, or question, and three response options, recorded as A, B, or C. Two item writer training sessions on writing performance-type, multiple-choice items for ***NCEXTEND1*** assessments were facilitated by NCSU-TOPS staff in April and June 2011. The first two-day session was held from April 4-5, 2011, and a second session was held from June 16-17, 2011. *Figure 3.6* below shows a template on writing ***NCEXTEND1*** items that was used in the item writer training (see Appendix 3-C).

Figure 3.6 NCEXTEND1 *Performance-based multiple-choice template*



About 207 teachers and educators from across the state took part in these sessions. *Table 3.6* shows a summary breakdown of teachers who participated in item training sessions. Following the training sessions in April of 2011, teachers were invited back in June of 2011 to participate in an active item writing workshop. Specific emphasis was placed on teachers who had experience with serving students with severe cognitive disabilities. The item writers were selected with diverse backgrounds based on their knowledge of the current Extended Content Standards, gender, ethnicity, and region. The use of North Carolina educators to develop items and their experience with the Extended Content Standards strengthened the face validity of the process.

Table 3.6 ***NCEXTEND1*** *Item Writing Training by Grade and Content Area*

Type	ELA			Math			Science		
	3-5	6-8	English II	3-5	6-8	Math I	5	8	Biology
General Education	10	8	8	8	7	7	7	3	3
Exceptional Children	21	20	16	20	18	11	15	16	9

3.3.4 Item Alignment

A critical aspect of item quality is alignment. Alignment refers to the extent to which an item agrees with and represents the content standard it is designed to measure. Assessments composed of items that are misaligned will generate scores that do not measure the breadth and depth of the intended construct. Scores from a misaligned assessment are characterized with high construct irrelevance variance and will underestimate or overestimate students' achievement. For this reason, alignment evidence is one of the most important sources of content validity.

During the item development phase, two groups were responsible for item alignment: 1) content specialists at the NCSU-TOPS and 2) members of the NCDPI/Curriculum and Instruction Division[3]. These groups independently reviewed proposed items and classified them by the Extended Content Standards and DOK levels. If any group felt that the written item did not classify exactly into the content standards and DOK, the item was revised until it aligned. A detailed report of the Alignment study is presented in Chapter 10.

3.4 Step 9. Review of Field Test Items

To ensure that items developed were aligned to the Extended Content Standards in ELA, math, and science, each item went through a detailed review process before being placed on a field test form. The Standard 3.2 of the *Standards* (AERA, APA, & NCME, 2014) states:

> *"Test developers are responsible for developing tests that measure the intended construct and for minimizing the potential for tests' being affected by construct-irrelevant*

[3]The NCDPI/test development created an alignment plan in 2010 before the development of any items. The alignment plan was reviewed by an expert in content alignment, Dr. Karen Hess, from the National Center for the Improvement of Educational Assessment. Based on her recommendations, an alignment plan was devised that would pre-align test items to the NC content standards.

characteristics, such as linguistic, communicative, cognitive, cultural, physical, or other characteristics."

A subset of teachers who participated in item writer training sessions were recruited and invited to review ***NCEXTEND1*** items during review sessions on September 12 and October 7, 2011. All items were reviewed by both general education teachers and special education teachers. The focus of the general education teachers was to assure that the academic content of the items was accurate and correctly linked to the appropriate grade-level extended content standard. Special education teachers reviewed each item for appropriateness and meaningfulness for the targeted population, availability of materials, clarity of directions, and accessibility. Based on the comments from the reviewers, items were revised and/or rewritten, item matching to an objective was reexamined and changed when necessary, and introductions and diagrams for items were refined. Additional items were developed as necessary to ensure sufficiency of the item pool. At the final step, test development staff members, with input from the curriculum staff and other content, curriculum, and testing experts, approved each item. The criteria for evaluating each written item included the following:

1. Conceptual

- Objective match (curricular appropriateness)
- Webb's Depth-of-Knowledge match
- Fair representation
- Lack of bias or sensitivity
- Clear statement
- 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/Manipulative Cards

- Necessary
- Plain (fewer details, simplify)
- Relevant
- Unbiased
- No shading within an object (e.g., Dog: no spots, keep solid)

3.5 Steps 10–11: Assembling and Reviewing Field Test Forms

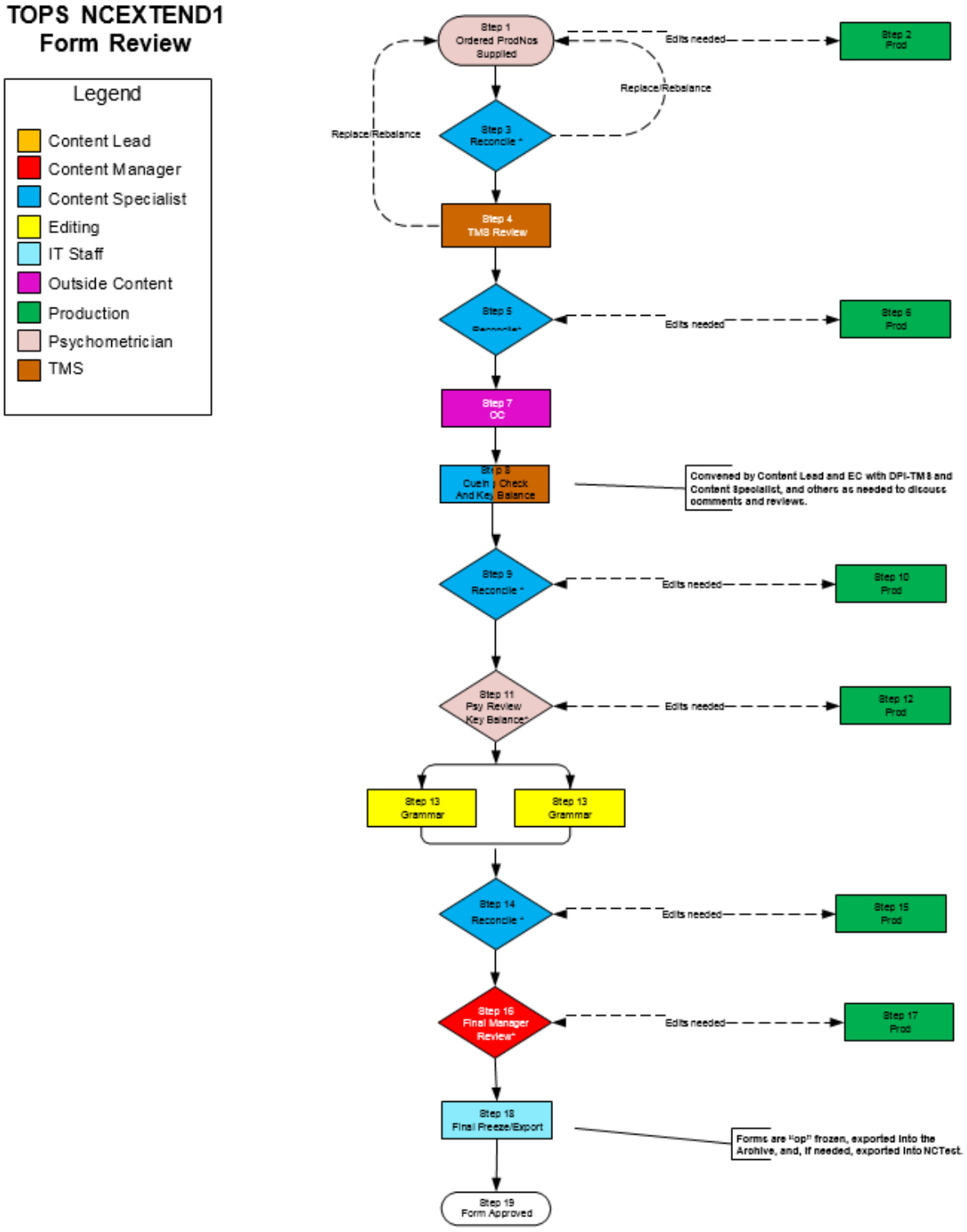
Field test forms were assembled to match the approved content specifications and blueprints following the NCDPI's 19-step iterative form building and review process for ***NCEXTEND1*** alternate assessments (see Figure 3.7). Field test forms were built according to the blueprints that were established during the standard prioritization meetings. For ***NCEXTEND1*** field test forms, the form building process was modified to address the needs of the ***NCEXTEND1*** population. Exceptional Children and Special Education experts at NCSU-TOPS and the NCDPI/Curriculum and Instruction Division played central roles in form assembly. The role of the test measurement specialists was to make sure that the test forms are aligned to the blueprints and most importantly to ensure their appropriateness for the ***NCEXTEND1*** population. The focus was to make sure combinations of items with their respective manipulatives and cognitive expectation were balanced. A subset of teachers who participated in item writer training sessions were recruited and invited to review the ***NCEXTEND1*** assessment forms on October 28, 2011. All forms were reviewed by both general education teachers and special education teachers.

Table 3.7 shows the number of forms, number of items in each form, and total number of items administered in the 2011–2012 stand-alone field test.

Table 3.7 Number of **NCEXTEND1** *Items Field Tested in Spring 2012*

Grade	Content	Number of Forms by Content	Number of Items per Form	Total Number of Items Field Tested by Content
3	ELA/Math	3	15	45
4	ELA/Math	3	15	45
5	ELA/Math/Science	3	15	45
6	ELA/Math	3	15	45
7	ELA/Math	3	15	45
8	ELA/Math/Science	3	15	45
10	English II/Math I /Biology	3	15	45

Figure 3.7 NCEXTEND1 Form Review Steps



Chapter 4 Field Test Administration and Operational Form Construction

The ***NCEXTEND1*** stand-alone field test was administered in Spring 2012. This chapter describes the field test administration and analysis of data collected during audits that were conducted in conjunction with the stand-alone field test. The final section of this chapter summarizes the final steps that were implemented to create the final operational ***NCEXTEND1*** forms.

4.1 Step 12: Field Test Sample and Administration[4]

The ***NCEXTEND1*** field tests were administered in February and March of 2012 to all eligible students enrolled in grades 3–8 and 10. The ***NCEXTEND1*** subjects were randomly assigned to students within class in such a way that each student was only administered one subject. For example, if there were four ***NCEXTEND1*** students in a grade 3 class, two were randomly administered the ELA/reading field test and the other two were administered the math. Each student was assessed in one-on-one basis where an Assessor read aloud all items and answer options with the exception of the last selection on the ELA/reading field tests. A proctor supervises the process, making sure the *Assessment Guide* is strictly followed.

Summary of ***NCEXTEND1*** students who participated in the field test by major demographic variables and subject are shown in *Table 4.1* for grades 3–8 and in *Table 4.2* for grade 10. Demographic distribution of the field test sample in each subject and population across the respective grades show a similar pattern across the major demographic variables. There is about 2:1 ratio of males to females in the ***NCEXTEND1*** population and this same trend is consistent across all grades and subject. Overall, the field test sample for each subject is representative of ***NCEXTEND1*** student population at the respective grade levels, and sample statistics can be generalized and interpreted to reflect population parameters with reasonable

[4] The NCDPI employs the same administration procedures for the field test and the operational assessment. Please see Chapter 5 for a detailed discussion of North Carolina's administration procedures.

levels of sampling error. The methods and results comply with Standard 1.8 of the AERA, APA, & NCME (2014) Standards, which states:

> *"The composition of any sample of test takers from which validity evidence is obtained should be described in as much detail as is practical and permissible, including major relevant socio-demographic and developmental characteristics."* (p. 25).

Table 4.1 Demographic Summary of Field Test 2012 Participants, NCEXTEND1 *Grades 3–8*

Grade	Category	N	Gender (%)		Ethnicity (%)							Special Subgroup (%)	
			Female	Male	Asian	Black	Hispanic	American Indian	Multiracial	Native Hawaiian/Pacific Islander	White	EL[5]	EDS[6]
3	Population	1,054	32.5	67.5	2.8	34.9	12.1	2.2	3.8		44.3	7.7	70.6
	Math	441	30.4	69.6	3.2	34.2	12.2	2.7	3.6		44.0	8.4	73.0
	ELA	444	32.4	67.6	1.8	34.7	13.5	1.6	3.6		44.8	7.2	67.8
4	Population	1,064	30.6	69.5	1.2	32.6	11.2	1.3	4.3		49.3	6.6	66.1
	Math	476	32.4	67.7	2.1	32.8	10.9	0.8	4.8		48.5	8.2	66.4
	ELA	445	29.9	70.1	1.1	34.4	10.8	1.8	3.8		48.1	4.9	64.5
5	Population	1,163	33.8	66.2	2.4	34.9	9.9	1.7	3.0	0.1	48.0	6.4	65.7
	Math	335	30.5	69.6	2.1	37.0	9.6	1.5	2.7	0.6	46.6	4.2	63.6
	ELA	342	35.1	64.9	2.6	32.5	7.9	1.5	3.5		52.1	6.1	64.9
	Science	325	36.0	64.0	1.9	35.7	12.0	0.9	3.1		46.5	6.5	68.6
6	Population	1,042	36.4	63.6	2.2	35.4	9.8	3.2	2.4	0.1	46.9	6.1	68.0
	Math	435	36.6	63.5	2.3	33.6	9.7	3.5	2.3	0.2	48.5	5.5	70.3
	ELA	435	34.3	65.8	1.6	39.3	10.3	2.1	2.3		44.4	6.7	69.2
7	Population	1,021	32.6	67.4	2.1	36.1	9.4	1.6	2.3		48.6	5.0	68.6
	Math	443	33.9	66.1	2.3	36.1	9.0	1.8	1.6		49.2	4.1	70.0
	ELA	462	33.3	66.7	1.7	37.9	8.7	1.5	2.4		47.8	5.2	67.1
8	Population	1,085	35.0	65.0	1.1	38.5	9.2	1.8	3.0	0.1	46.2	4.0	68.2
	Math	341	36.1	63.9	0.9	40.5	10.0	1.8	1.8		45.2	4.1	66.9
	ELA	333	34.5	65.5	1.5	38.1	8.4	2.1	2.4		47.5	3.6	68.5
	Science	300	34.7	65.3	0.3	39.3	10.7	1.3	3.7		44.7	3.3	69.3

[5] English Learner
[6] Economically Disadvantages Students

Table 4.2 Demographic Summary of Field Test 2012 Participants, NCEXTEND1 *Grade 10*

Grade	Category	N	Gender (%)		Ethnicity (%)							Special Subgroup (%)	
			Female	Male	Asian	Black	Hispanic	American Indian	Multiracial	Native Hawaiian/Pacific Islander	White	EL[7]	EDS[8]
10	Population	786	35.1	64.9	1.9	37.2	9.2	0.9	2.7		48.2	2.9	62.5
	Math I	223	36.8	63.2	1.8	36.3	7.6	0.5	2.7		51.1	1.8	64.1
	ELA II	224	32.1	67.9	1.8	40.2	9.8		3.1		45.1	1.8	63.4
	Biology	225	38.7	61.3	2.7	36.9	8.9	1.3	3.1		47.1	3.6	62.7

[7] English Learner
[8] Economically Disadvantages Students

4.2 Step 13. Field Test Analyses

Field test data analyses provided evidence used to determine whether items performed as expected and were suitable for use on operational forms. Edition 3 of the ***NCEXTEND1*** assessment was redesigned with new performance-based multiple-choice items that used visual display items and answer cards and also with new administration protocols for Assessors. Analyses of field test data were vital in providing evidence as to whether items and test forms perform as expected.

4.2.1 Audit Report from 2011–2012 Field Test

As part of the field test analysis on the redesign of ***NCEXTEND1***, on-site audits were conducted at selected schools with the goals to verify appropriate implementation of ***NCEXTEND1*** eligibility criteria and also to evaluate the implementation of the newly designed assessments. These audits were conducted February 27, 2012, to March 23, 2012, to coincide with the ***NCEXTEND1*** field test administration window in six schools throughout the state across grades 3–8, and 10. The audits focused on three specific areas: (1) the level of adherence to test administration and scoring procedures, (2) the use of accommodations, and (3) the level of compliance with the ***NCEXTEND1*** eligibility criteria. A combination of qualitative and quantitative research methodologies was used to prepare for and conduct the 2011–2012 ***NCEXTEND1***audits.

These next sections summarize the entire field test audit process as captured in the full audit report that was prepared following synthesis from all the separate audits. Separate audit reports by school were also prepared and presented to each LEA highlighting specific findings and recommended actions needed as noted by auditors. The findings of the audit were intended to provide useful information for improving the monitoring of eligible students, quality and administration of the ***NCEXTEND1*** Edition 3 assessments**.**

- **Sample**

A purposeful sampling of LEAs, charter schools, and alternative special education schools was conducted in August 2011. Two schools were selected as a result of data review

during the annual 1% cap waivers meeting. Review of historical data showed that these two schools had a large number of students participating in the previous editions of the ***NCEXTEND1*** assessment with unusually high performance at the lower grades. Two other schools were selected as a result of calls the NCDPI received from test administrators regarding potential concerns with the inappropriate use of testing accommodations. The primary concern involved teachers augmenting the independent reading selections by substituting words for picture symbols. The final two schools were selected at random.

This selection process was designed to guarantee equal representation from all ethnic, gender, and socioeconomic subgroups across all regions of the state. A listing of the LEAs, grade levels, regions, and number of audits conducted is provided in *Table 4.3*.

Table 4.3 2011–2012 Audit Sample

LEA	Grade Level/Type	Region	Number of ***NCEXTEND1*** Students
Bertie	Elementary	Northeast	4
Charlotte-Mecklenburg Schools (2)	Middle K–12 Special Education	Southwest	19
Lee	K–12 Special Education	Central	12
Martin	Elementary	Northeast	2
Vance	Middle	Central	10

A total of 44 students across all assessed grade levels were selected for observation. A listing of the number of students per grade level is provided in *Table 4.4*.

Table 4.4 Audit Sample ***NCEXTEND1***—Number of Students by Grade Level

Grade Level	Number of Students
Grade 3	3
Grade 4	8
Grade 5	4
Grade 6	8
Grade 7	6
Grade 8	12
Grade 10	3

Schools were notified of their selection early in the academic year and were provided a description of what their participation entailed during a series of conference calls with LEA- and school-level testing personnel.

- **The Auditors and Training**

Audit team members, representing NCDPI Test Development, Testing Policy and Operations, Exceptional Children, Curriculum and Instruction, and NCSU-TOPS were trained on the audit process and all data collection procedures. Auditors were also trained on characteristics/behaviors of the student population, use of accommodations, and possible student response modes (e.g. eye gaze, finger pointing, and use of switches). These training and informational processes helped to ensure the integrity and standardization of the assessment were maintained and valid inferences could be made from data collected during the audit. A complete listing of all auditors is provided in *Table 4.5*.

Table 4.5 2011–2012 Auditors

Auditor Area of Responsibility	Count
NCDPI–Testing Policy and Operations	2
NCDPI–Test Development	4
NCSU–TOPS	1
NCDPI–Accountability Services	1
NCDPI–Exceptional Children Division	1
NCDPI–Curriculum and Instruction Division	1

- **Audit Process and Procedures**

Upon arrival for each observation, auditors reported to the school's administrative office where they were greeted by the school's testing coordinator. After the auditors presented proper identification and signed the school's visitor log, the testing coordinator then directed them to the test administration location. Some administrations were conducted in the student's classroom while others were conducted in conference rooms or unused classrooms. For each test administration, the auditors consulted with the student's teacher, usually the Assessor, to determine how best to introduce themselves to the student, the testing environment, and where the auditors should sit in relationship to the student. This information was needed to ensure the auditor's presence in the classroom would not have any negative effect on the student's performance.

During the test administration, the auditors read the test booklets and reading selections following along as the Assessor read to the student and recorded whether the Assessors were appropriately following the script. The auditors also recorded the student's answer choice for each item and the time length of the test administration. Auditors were provided with a copy of the student's test record and Individualized Education Program (IEP) found in the student's cumulative folder. The use of accommodations and/or modifications observed during the testing administration was also recorded and compared to the information found in the student's IEP. Copies of the student's test record and IEP were submitted to the NCDPI along with the other data collection sheets for further analysis.

Before leaving each school, the auditors were required to sign the student's confidential folder to document their review of the student's IEP and other confidential information. After the auditors had completed all forms and gathered all required documentation, they were directed back to the school's administrative office where they signed out and left.

- **Results and Findings**

The ***NCEXTEND1*** field tests were administered from February 27, 2012, to March 23, 2012. While 44 students were selected for participation in the ***NCEXTEND1*** audit, a total of 42 complete test administrations were observed. Two field test administrations were stopped as a

result of student behavior. A listing of all observed test administrations by grade and content area is provided in *Table 4.6*.

Table 4.6 2011–12 Audit Sample—Observed Test Administrations by Grade and Subject

Grade/Subject	Number of Tests Observed
Grade 3 Reading	1
Grade 3 Mathematics	2
Grade 4 Reading	3
Grade 4 Mathematics	4
Grade 5 Reading	2
Grade 5 Mathematics	2
Grade 6 Reading	3
Grade 6 Mathematics	4
Grade 7 Reading	3
Grade 7 Mathematics	3
Grade 8 Reading	5
Grade 8 Mathematics	2
Grade 8 Science	5
Grade 10 Mathematics	1
Grade 10 Science	2

- **Assessors' Demographics**

The ***NCEXTEND1*** *Assessment Guide* stipulates that Assessors must have professional training in education and the testing program. More specifically, the person identified as the

Assessor must have routine contact with the student during classroom instruction and must be the student's primary teacher for the assigned content area.

Twenty-nine special educators, assigned as either Assessor or Proctor, were observed during the 42 test administrations. The role of the Proctor was to supervise the process for appropriate and secure test administration. An Assessor for one student may act as a Proctor for other student and vice versa. The Assessor population consisted primarily of Exceptional Children teachers. Some of the Proctors were identified as teaching assistants. Review of the data collected regarding the Assessors showed that 14 Assessors have bachelor's degrees and 8 Assessors have master's degrees. In addition, 66% of the Assessors/Proctors (19) have at least five years of teaching experience. A detailed listing of the Assessors/Proctors' years of experience is provided in *Table 4.7*.

Table 4.7 2011–2012 Audit—Assessors' Years of Teaching Experience

Years of Teaching Experience	Number of Teachers
0 – 4	10
5 – 10	7
11 – 15	9
16 – 20	0
21 or more	3

Following each test administration, Assessors were asked to provide the auditors with feedback regarding test administration procedures and potential training needs. This information was collected and was included in the training processes for the operational field-test administration of the ***NCEXTEND1*** assessments in the 2012–2013 academic year.

- **Finding 1: Adherence to Test Administration and Scoring Procedures**

Edition 3 of the ***NCEXTEND1*** is marked by several changes to the assessment's design, development, administration, and scoring processes. All items were developed by current

teachers of students with significant disabilities in collaboration with regular education teachers. All test items followed a multiple-choice format with a single stem, or question, and three response options, recorded as A, B, or C. During the development process, all test forms, consisting of 15 items, were later reviewed by these same teachers to assess the appropriateness and usability of the items, the item scripts, and assessment directions.

The most significant change to the ***NCEXTEND1*** assessment design involved allowing students up to two attempts for each item, otherwise referred to as the "2-Trial Method." The ***NCEXTEND1*** Edition 2 assessment only allowed one attempt for each item. For situations when students are nonresponsive or no clear choice is made, Assessors may re-present the item up to two additional times.

The guidance around what was considered "nonresponsive" in Edition 2 presented several questions for the test administrators given the variability of student behaviors and characteristics across the assessed population. As a result, beginning with Edition 3 of the ***NCEXTEND1*** assessment, students are allowed up to two attempts to select a response for each item. On the first attempt, students are prompted to select a single response. If the student selects the correct response option, the item scripts in the Assessor Booklet provide directions for continuing to the next item. However, if the student selects an incorrect response option, makes no clear response, or is nonresponsive, the incorrect answer choice or an identified answer choice is removed and the item is re-presented a second time with only two response options. Conversations with teachers of the assessed student population revealed that presenting items with two response options more closely mimicked regular classroom instructional and assessment practices for many students where a concentrated amount of time is spent on making choices between two objects or ideas.

These changes to the administration process also had implications for scoring. In the previous edition of the test, students received two score points, one from each Assessor, for each correct response and zero points for each incorrect response. The only way students could earn an odd number of score points was if the Assessors disagreed as to whether the student selected the correct answer or not. In addition, because actual student responses (A, B, C) are now being recorded in the current edition as opposed to "Yes/No," in the previous edition of the

NCEXTEND1 tests, the number of Assessors for the current ***NCEXTEND1*** was reduced to one, and the second Assessor now assumes the role of a proctor. With the new assessment design, the student can earn two score points for items that are answered correctly on the first attempt and one score point for items that are answered correctly on the second attempt.

Given these changes to the assessment's design, auditors collected data on student responses, the length of the administration, and the impact of the 2-Trial Method on raw score values. The auditors also provided additional comments regarding the accessibility of test materials for students of varying areas of disability and the detection of bias in the items and the reading selections across test forms.

- **Implications for the 2-Trial Method**

A sub goal of the audit was to ensure that the implementation of the 2-Trial Method did not result in a negative impact of student performance in terms of length of administration and a reduction in the student's ability to access the assessment. Prior to the ***NCEXTEND1*** Edition 3 stand-alone field test audit, only a limited amount of information regarding the expected time for a single test administration had been provided. Review of time data for the audited student population showed an average of 26 minutes for each field test administration and an overall range of 14 to 59 minutes across the 42 students. *Table 4.8* below denotes the average time lengths across subject areas and *Table 4.9* denotes the average time length for the assessments across eligibility areas.

Table 4.8 Average Time Length for NCXTEND1 Assessment across Subjects

Subject	Minutes for Assessment
English language arts (ELA)	28
Mathematics	27
Science	17

Table 4.9 Average Time Length for **NCEXTEND1** *Assessment across Eligibility Areas*

Eligibility Area	Minutes for Assessment
Intellectually Disabled Mild (IDMI)	26
Intellectually Disabled Moderate (IDMO)	29
Intellectually Disabled Severe (IDSE)	31
Autistic (AU)	24
Multiple Disabilities (MU)	23

Time for testing reflects the actual minutes students spent responding to items. Minutes for breaks during the test administration are not included. Review of student response data also showed that the 2-Trial Method increased the number of raw-score points students were able to earn. The students were able to correctly respond to on average five additional items using the second response trial. As a result, these additional correct responses increased the raw-score values for these students by an average of five points resulting in total raw scores of 10 to 26 points.

- **Adherence to Directions and Scripts**

All tests that are part of the North Carolina Testing Program require a standardized process of administration. For the test results to be valid, all procedures outlined in the ***NCEXTEND1*** *Assessment Guide* must be followed. Assessors are responsible for following all directions in the Assessment Guide and all assessment booklets. As indicated in the Assessment Guide, paraphrasing, omitting, revising, or rewriting the scripts or the directions contained within the Assessor Booklet is considered a testing irregularity. Additionally, Assessors are not allowed to confer regarding student responses to test items. Failure to comply with any of these directions is a direct violation of the North Carolina *Testing Code of Ethics*.

Review of collected audit data across all schools showed that Assessors followed most test administration directions and scripts. Auditors across the entire audit sample commonly

noted careful attention paid to item presentation, maintenance of student engagement throughout the assessment, and evidence of appropriate rapport between Assessors and students.

Deeper analysis of audit data showed Assessors followed all scripts and directions in 78.3% (33 out of 42) of the observed test administrations. The most frequently noted irregularity included the omission of break scripts for students who took breaks during the test administration and the omission of the "End of Assessment" statement following the presentation of the last item on the test form. In two of these test administrations at one of the audited schools, the auditors noted where the Assessor prompted the student when the calculator was needed before responding to items on two separate mathematics assessments. Information regarding these occurrences was shared with the LEA and school principal during the school's exit conference call and safeguard to avoid future occurrences was implemented into the operational ***NCEXTEND1*** training processes.

Other minor issues regarding mix-ups with presentation of response cards and presenting the second trial for some items when it was not needed were also noted. However, these events were not noted as irregularities for the purposes of the field test audit. The audit staff attributed these occurrences to the newness of the assessment design, changes in the administration processes, and the level of comfort of the Assessors.

- **Considerations for NCDPI Training and Continuous Improvement**

Review of the field test audit data highlighted several potential training needs that the NCDPI incorporated into the ***NCEXTEND1*** training processes for the 2012–2013 academic year. First, the NCDPI developed sample ***NCEXTEND1*** items that closely represented those test items used on the operational test forms across subjects and grade levels. These items were available for public use and posted on the NCDPI website.

Second, the NCDPI developed sample ***NCEXTEND1*** student test administration videos. These videos assisted LEA test coordinators and school-based testing coordinators in developing a more authentic training process for Assessors. These videos included students from different areas of disability, grade levels, and content areas. Given the nature of this student population,

these videos are maintained for training purposes only and are not posted for general public use on the NCDPI website.

Third, stemming from conversations with the observed Assessors, the NCDPI made adjustments to the ***NCEXTEND1*** Assessment Guide and Assessor Booklets so that the "Directions for the Test Administration," "Introduction of the Assessment" scripts, "Breaks during the Testing Session" scripts, and "End of Assessment" scripts could be incorporated into initial ***NCEXTEND1*** Assessor training processes. In addition, the Assessment Guide and the Assessor Booklet included suggestions or strategies to assist with the removal of identified response cards and the recording of student responses for the 2-Trial Method.

Based on these audits, the NCDPI also revisited ***NCEXTEND1*** policies regarding the previewing of secure testing materials. Given the number of changes to the ***NCEXTEND1*** test design, decisions were made regarding the appropriate time frame in which Assessors can preview the secure test materials prior to the student test administration. These policy changes were included in the ***NCEXTEND1*** *Assessment Guide* and the *Testing Students with Disabilities* document.

These proposed changes to the training processes for ***NCEXTEND1*** test administration should 1) increase familiarity with the new ***NCEXTEND1*** assessment design, 2) reduce anxiety for both Assessors and students, and 3) improve the overall flow and pace of the test administration for the operational test administration in 2012–13 and beyond.

- **Finding 2: Review of Accommodations**

Following the plain language principles, all items for the ***NCEXTEND1*** assessments were created to ensure maximum accessibility for all students. However, on a case-by-case basis, accommodations may be made for individual students for both presenting and/or responding if the accommodation is routinely used during instruction and other classroom assessments. A list of approved accommodations for the ***NCEXTEND1*** tests of ELA, mathematics, and science can be found in the Assessment Guide. Data and information regarding the use of testing accommodations was collected for each student observed. In particular, the auditors looked for a

direct linkage between the testing accommodations used during the actual test administration with those documented in each student's IEP.

Given the design of the ***NCEXTEND1*** test administration, certain testing accommodations are required by the ***NCEXTEND1*** test administration procedures and were observed in all test administrations. These accommodations included the following: separate setting, test administrator reads test aloud in English, extended time, and dictation to scribe. Also embedded in the test design are the break scripts found on page 2 of the Assessor Booklet. These allow for students to take as many breaks as needed during the test administration in order to maintain engagement and ensure optimal student performance.

Adjustments were made to the collected audit data to exclude those accommodations required by the test administration procedures. Further review of students' IEPs documented the provision of unique, student-specific testing accommodations for 28.3% (13 out of 46) of the observed field test administration.

Table 4.10 provides a detailed listing of testing accommodations documented in student IEPs and those that were not observed/used during the audit.

Table 4.10 2011–2012 Audit—Documented vs. Not Observed Testing Accommodations

Testing Accommodation	Documented	Number (*N*) Not Observed	Percentage of Use
More Frequent Breaks	2	2	0%
Adaptations to NCDPI-provided Manipulative Cards	10	1	90%
Multiple Testing Sessions	10	4	60%
Augmentative Communication Devices	5	4	20%
Assistive Technology	5	2	60%

Large Print Materials	5	5	0%

Review of audit data showed some inconsistency between those accommodations that were documented in individual student's IEP documentation and those that were actually observed. The most consistently documented and used external accommodations in IEPs included "Adaptations to the NCDPI-provided Manipulative Cards" and "Multiple Testing Sessions." During test, though, the provision of "Large Print Materials" showed the lowest rate of use. This external accommodation is typically reserved for students whose disabilities require the administrator to increase the size of the provided test materials beyond the printed 22-pt font size. The absence of this provision suggests that this accommodation is not needed for these students and should be reviewed during these students' respective annual IEP review meetings.

The provision of "More Frequent Breaks" also showed an equally low rate of use; however, further analysis of IEP documentation showed that the allowed time intervals for breaks were greater than the length of time it took for these two individual students to complete the field test administrations. This finding was also shared with the schools so that potential adjustments could be made to each student's accommodations documentation.

These findings regarding the inconsistent use of external testing accommodations suggested the need for more training and information regarding the use of breaks embedded in the test design and the administration procedures.

- **Considerations for NCDPI Training and Continuous Improvement**

Starting with the 2012–2013 academic year, accommodations that were considered to be either embedded in the test design or not applicable to the ***NCEXTEND1*** assessments were removed from the ***NCEXTEND1*** Approved Accommodations chart. These accommodations include "test administrator reads test aloud in English," "dictation to scribe," "Braille writer," and "student reads test aloud to self." The removal of these accommodations may reduce the level of questions IEP teams may have regarding which accommodations should be used and documented. The NCDPI also planned to provide continued training on the appropriate use of "multiple testing sessions" and "breaks during the assessment." This training, along with the

time for test administration collected, should aid test coordinators and Assessors in scheduling and planning during the ***NCEXTEND1*** testing window. The NCDPI has also proposed to provide electronic resources to aid Assessors in the appropriate use of "adaptations to the NCDPI-provided manipulative cards," and to post these resources on the NCDPI website and become part of the ***NCEXTEND1*** training materials at the beginning of the 2012–2013 academic year.

Lastly, the NCDPI planned to publish an "***NCEXTEND1*** Graphics Pictionary" that documents the most frequently used graphics and pictures across all ***NCEXTEND1*** manipulative cards. This Pictionary assists Assessors in understanding how to appropriately augment the NCDPI provided manipulative cards if needed. Additionally, this Pictionary provides Assessors with the ability to incorporate these graphics into regular classroom instruction and assessment through the school year. The "***NCEXTEND1*** Graphics Pictionary" was available for general public use and housed on the NCDPI website.

- **Finding 3: Compliance with *NCEXTEND1* Eligibility Criteria**

The ***NCEXTEND1*** eligibility criteria requires that students have an IEP and have a significant cognitive disability (i.e., exhibit 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). Students participating in the ***NCEXTEND1*** assessment must also be instructed on the North Carolina Extended Content Standards for their assigned grade level. Any instruction of general education content, even at a lower grade level, is *NOT* appropriate for students participating in this assessment. Each student's IEP was reviewed to 1) determine their area of disability, 2) document evidence of appropriate academic instruction as indicated by IEP goals, and 3) review alternate assessment justification statements.

The Individuals with Disabilities Education Act of 2001 (IDEA) currently recognizes 13 different areas of disability for students in public education. The students selected to participate in the audit included students across all disability/eligibility areas. Some eligibility areas such as deaf-blindness, serious emotional disability, hearing impairment, and visual impairment were not represented in the ***NCEXTEND1*** sample because these were not the primary eligibility areas of

the selected students. Some students assessed using the ***NCEXTEND1*** assessment may have secondary areas of eligibility, which may include these disability/eligibility areas. *Table 4.11* provides a listing of the primary areas of disability for the entire student population.

Table 4.11 2011–2012 Field Test Audit—Sample Student Population by Areas of Disability

Disability Code	Disability Description	Number of Students
AU	Autistic	20
ID-MI	Intellectually Disabled—Mild	4
ID-MO	Intellectually Disabled—Moderate	15
ID-SE	Intellectually Disabled—Severe	4
MU	Multiple Disabilities	3

- **Instruction on the North Carolina Extended Content Standards**

The ***NCEXTEND1*** is designed to assess student understanding of English language arts, mathematics, and science content outlined in the North Carolina Extended *Common Core State Standards* for ELA and mathematics and the North Carolina Extended *Essential Standards for Science* in each assessed grade level. Therefore, each student's IEP must also reflect at least one academic goal for ELA, mathematics, and science. Review of student IEP documentation for the selected student population showed 100% of students are receiving instruction on the North Carolina Extended Content Standards at their assigned grade level.

- **Alternate Assessment Justification Statements**

The alternate assessment justification statement in the student's IEP requires IEP teams to provide a rationale for why the regular testing program, with accommodations, is not appropriate and why the alternate assessment, with or without accommodations, is appropriate. Therefore, the justification statement for a student identified as Intellectually Disabled—Mild (IMDI) should not be the same as a student who is identified as Intellectually Disabled—Severe (IDSE), as their skills and abilities are extremely different from each other. Review of students' IEPs showed that 100% of students' IEPs provided an alternate assessment justification statement. Most of the students' justification statements noted significant deficits in academic ability and

the presence of significantly low cognitive abilities and were written to reflect the learning needs of individual students.

- **Review of Students' Test Records**

Each student's test record was reviewed in order to ensure the student was appropriately categorized to be in the ***NCEXTEND1*** population. Several student test records indicated participation on the general assessment and/or the ***NCEXTEND2*** modified assessment at previous grade levels. Both of these assessments are intended to assess student understanding of the North Carolina *Standard Course of Study* general education curriculum. Typically, student's scores indicated proficiency on the North *Carolina Standard Course of Study* at the end of one year but not proficient the next year, whether the student was administered the general assessment or the ***NCEXTEND2***. Special concerns were noted in the audit data when individual students were identified as requiring the ***NCEXTEND1*** *and* had demonstrated academic growth or even proficiency on the general education curriculum and corresponding assessments. These concerns were discussed with each LEA to ensure they are monitoring the appropriate use of the assessments with the focus being on administering assessments that are aligned to students' instruction and academic ability.

- **Special Concerns Regarding Adherence to *NCEXTEND1* Eligibility Criteria**

Further review of IEP documentation, student test records, and student performance highlighted some concerns regarding assessment eligibility for seven students at four different schools. As previously discussed, two of the audited schools were selected for participation in the ***NCEXTEND1*** field test audit as a result of 1% cap waiver requests and review of historical ***NCEXTEND1*** test scores where high numbers of students across the LEA were earning unusually high test scores without any evidence of prior testing on the regular education assessments. Two other schools were selected as result of concerns regarding the inappropriate augmentation of testing materials and the final two schools were randomly selected.

Three of the seven students were from the two purposefully selected schools from the 1% cap waiver review. Thorough review of the audit data showed evidence of consistency with the concerns that were noted during the 1% cap waiver review. Observations of these students,

including test performance and individual conversations with audit staff raised some questions regarding the appropriateness of their participation in the alternate assessment program. As defined in the ***NCEXTEND1*** Eligibility Criteria, a student with a significant cognitive disability has severe and pervasive delays in ALL areas of conceptual, linguistic and academic development.

Two other students were from one of the other two schools where inappropriate accommodations were observed. Review of test performance and IEP documentation showed evidence of academic ability beyond the intent of the NC Extended Content Standards. These concerns were also noted by the classroom teacher of these two students, who readily discussed an upcoming re-evaluation and the potential movement of these students to both a regular education setting and the ***NCEXTEND2*** assessment. Information regarding the alignment of assessment participation to the instruction received was discussed with the schools and LEA staff. All LEAs and schools were reminded that a change in the student's instruction did not always warrant reclassification of students to the least restrictive environment. The decisions surrounding instruction and learning should remain separate.

The two remaining students were from one of the two randomly selected schools. Like the other students, these two students also showed evidence of academic ability beyond the intent of the Extended Content Standards. Similar questions regarding the alignment between each student's least restrictive environment and the type of instruction received were also noted by the school and LEA staff. As with the previous school, this school was also reminded of the need to align assessment decisions to instruction and academic ability as opposed to the student's learning environment.

Each of the seven students across these four schools showed evidence of relatively high functionality, linguistics, and academic ability beyond that which is intended for the population of students assessed using the ***NCEXTEND1***. The cognitive functioning of these seven students seemed to more accurately reflect the definition of a specific learning disability as opposed to significant cognitive disabilities. These findings were shared with the schools and LEA staff during each school's Audit Exit Conference Call. Each school was reminded of the long-term implications these decisions have on students in terms of high school graduation and post-

secondary opportunities. Continued training and information regarding the appropriate instructional and assessment decisions for this student population is needed and will continue to be included in future auditing of the ***NCEXTEND1*** assessment.

- **Conclusions**

After reviewing all information from the audit, the NCDPI recognized that more training and information, including administration procedures and eligibility requirement, regarding the ***NCEXTEND1***, must be shared with other NCDPI staff, LEA testing personnel, school administrators, and teachers. To ensure additional trainings occurred, the NCDPI had proposed to:

- continue auditing administrations of the ***NCEXTEND1*** assessments annually,
- provide participating LEAs with an official report of the findings as well as sharing general findings with all of the LEAs,
- provide additional training regarding the ***NCEXTEND1*** Assessment Guide,
- provide web-based training modules for off-site training,
- update information on the ***NCEXTEND1*** NCDPI Web page, and
- Continue communicating between NCDPI Test Development, Exceptional Children, and General Education staff regarding the ***NCEXTEND1***.

The expectation was that implementation of additional training measures would increase understanding of the processes for the operational ***NCEXTEND1*** Edition 3 administration, as well as increase awareness of the importance of making appropriate assessment choices for students that reflect their instruction and academic abilities. To ensure the continued awareness and effective practice of uniform and valid test processes for schools and IEP teams, the NCDPI shared all audit results and findings with the NCDPI Exceptional Children's Division and all other audit participants. This collaboration between the NCDPI Division of Accountability Services, Division of Exceptional Children, and Division of Curriculum and Instruction will continue to be an essential component of future auditing of the ***NCEXTEND1*** assessment.

4.2.2 Review of the Field Test Items

At the conclusion of the field test, NCSU-TOPS alternate assessment test development experts, in collaboration with the NCDPI Test Development, Curriculum and Instruction, and Exceptional Children sections, conducted reviews of all the audit feedback and teacher's item comments. The item comments were captured in the form of open-ended comments provided during field test administration. The results for each item were integrated in the NCDPI's online Test Development System, and provided additional evaluation of qualitative data from field test items. Following these in-depth analyses of all ***NCEXTEND1*** forms with associated data collected from field tests, the following major recommendations were reached:

1) Move selection to appropriate grade: in order to address the discrepancies between the selections' word count and readability across grade levels and within each grade level in ELA, expert reviewers decided to revise some selections from the field test, to move selections to the appropriate grade, and to include new selections when necessary. The final selections' readability at each grade is shown in *Table 4.12*.
2) Arrange items from easier to harder: The reviewers also decided to start the test with easiest selection, and the selections would increase in difficulty, except for the last selection that students read independently.
3) Reduce difficulty of the test: The group realized that some items were difficult for the 1% population and recommended to decrease difficulty of the overall form to meet the special needs of ***NCEXTEND1*** students.
4) Graphics: The group also recommended graphics be included in almost all the items.

Table 4.12 Readability Levels for Selections in **NCEXTEND1** *ELA/Reading*

Grade	Word Count	Flesch-Kincaid Reading Level
3	30–40	1.0–1.5
4	40–50	1.5–2.0
5	50–65	2.0–2.5
6	65–80	2.5–3.0
7	80–95	3.0–3.5

8	95–110	3.5–40
10	110–125	4.0–4.5

Based on all the above qualitative and quantitative review of ***NCEXTEND1*** tests, there was consensus that field test items needed significant revision and in some cases new items would have to be written. There was enough evidence from the audit report and teacher review that the current field test items will have to undergo significant revisions if they were to survive for operational use.

4.3 Step 15. *NCEXTEND1* Operational Field-Test Construction

Standard 3.2 of the *Standards* (AERA, APA, & NCME, 2014) states:
"Test developers are responsible for developing tests that measure the intended construct and for minimizing the potential for tests' being affected by construct-irrelevant characteristics, such as linguistic, communicative, cognitive, cultural, physical, or other characteristics." (p. 64).

As indicated in the previous section, most of the *NCEXTEND1* items in the stand-alone field test required significant revisions across all grades and content areas if they were to survive for operational use. Based on results from audit analysis and teacher feedback of the stand-alone field test administration of ***NCEXTEND1*** Edition 3, content experts at NCSU-TOPS, with support from NCDPI, agreed to deviate from their standard operational form building protocol. Exceptional Children staff from NCDPI and Special Education content experts from NCSU-TOPS were assigned the responsibilities to review and make appropriate revisions to all items and assemble a new form for all ***NCEXTEND1*** content areas and grade levels. The newly assembled form for each grade level and content area were administered in 2012–13 as ***NCEXTEND1*** Edition 3 operational field test. Test Measurement Specialists (TMS) and psychometricians assumed the roles of monitoring to ensure items and forms met test specifications and are aligned to NC extended standards. All the newly assembled forms went through the ***NCEXTEND1*** form review steps (see Appendix 4-A) for the complete form review

steps and narratives). Therefore, the 2012–13 administration of ***NCEXTEND1*** in the NC Statewide Testing Program was treated as an operational field-test. This allowed the opportunity for additional review of items before scores were certified and used for federal and school reports.

4.4 Operational Test Production

4.4.1 General Test

The ***NCEXTEND1*** materials consist of Assessor booklets, selection booklets for ELA, and manipulative card kits that contain the answer choices for each item. The following processes are followed when producing the operational tests:

1. Word files of the test booklets and manipulative cards are converted to PDF format by NCSU-TOPS production staff.
2. PDFs are reviewed by NCSU-TOPS Editing staff for the following:
    - formatting
    - grammar
    - readability
    - content
3. TOPS-Production prints booklets and manipulative cards and provides them to NCDPI-Operations for review.
4. NCDPI-Operations reviews hard copies of the materials. If errors are found, the document is sent back to step 1, edits are made, and steps 2 and 3 are repeated.
5. NCDPI-Operations approves documents for printing.
6. NCDPI-Production transfers files to a secure FTP site for pre-press processing.
7. The print vendor posts pre-press files for NCDPI-Operations to review.
8. The print vendor prints documents once NCDPI-Operations approves the pre-press files. If an error was found, the files are resubmitted and steps 6 and 7 are repeated.
9. Printed materials are shipped to the NCSU-TOPS warehouse.

10. Kits are assembled using test booklets, selection booklets, manipulative cards, and assessment guides.
11. Kits are shipped according to Assessors matched in NCEducation.

4.4.2 Braille Test

As in the general tests, the ***NCEXTEND1*** materials for Braille tests also consist of Assessor booklets, selection booklets for English Language Arts (ELA)/Reading, and manipulative card kits that contain the answer choices for each item. The Assessor booklets are not produced in Braille because the test administrator reads the student information aloud to the student. The following processes were followed when producing the operational tests:

1. ELA Selection Review

The NCDPI/Exceptional Children Division reviewed selection booklets for bias and accessibility for students with visual impairments and makes a recommendation either to use or reject a selection. Exceptional Children staff also created transcriber notes to describe illustrations used in the selections.

2. Item Review

The NCDPI/Exceptional Children Division reviewed items for bias and accessibility for students with visual impairments. They can make suggestions on ways to reword an item if imagery is used in its stem or foils. Additionally, graphics on the manipulative cards are evaluated upon their ability to be reproduced in Braille. If necessary, Exceptional Children staff make a recommendation not to use a graphic or suggest ways to improve its accessibility for Braille. The staff also created transcriber notes to describe graphics used in manipulatives that are not paired with text.

3. Form Review

The NCDPI/Exceptional Children staff reviewed the ELA selection booklets and manipulative card kits for their accessibility to Braille with particular attention to the accessibility of the graphics and the imagery used in the wording of the items. The following processes were followed when producing the Braille forms:

1. Editing staff posts the final PDF of the selection booklet and manipulative card kits on the NCDPI secure shell, which is accessible to the Braille staff.
2. The Braille staff sends proof copies of the labels that affix to the selection booklets and manipulative cards to NCSU-TOPS.
3. NCSU-TOPS assigns proof copies to certified Braille specialists (proof readers) for review.
4. Any proof notes received from the specialists are posted on the secure shell. The Braille staff make requested changes and produces the requested number of final Braille copies.
5. A second Braille specialist reviews one of the copies in the final series, comparing it to the proof copy/proof notes to ensure compliance.
6. If additional edits are needed, the senior editor requests corrected labels from the Braille staff.
7. Specialists assisted by trained in-house editors compare the remaining copies in the series to the form approved by the specialist to ensure they are all identical.
8. Braille labels are affixed to the reading selections and manipulative cards.

   After the forms were produced, the following processes were followed to ship forms and collect feedback:

   a. The final copies are labeled as Braille editions and sent to the warehouse for shipping.
   b. The NCSU-TOPS warehouse includes an error report form in each shipment for feedback.
   c. All error reports received are checked by Braille specialists, and changes are made by the Braille contractor, if necessary.

Chapter 5 Test Administration

This chapter of the technical report describes the materials and activities in which the NCDPI engaged in order to assure a uniform, fair administration of the test for all students across the state of North Carolina. The ***NCEXTEND1*** assessments are administered individually to students; in other words, an Assessor administers the ***NCEXTEND1*** to one student at a time. The Assessor is expected to follow an approved administration guide so all students with consideration of their special individual needs have a fair opportunity to demonstrate their understanding of the content being assessed so as to minimize construct-irrelevant variance that could undermine the comparability of test scores. The NCDPI produces an **NCEXTEND1** *Assessment Guide* which covers all aspects of test administrations to ensure test administration guidelines are adhered to so that score interpretation is valid.

5.1 *NCEXTEND1* Assessment Eligibility Criteria

The NCDPI provides guidelines and requirements for ***NCEXTEND1*** eligibility published in the *Assessment Guide*, which defines the population of students for which the assessment is designed. In order for any student to participate in the ***NCEXTEND1*** assessment, the IEP team should:

1. Determine that the student has a significant cognitive disability.
2. Determine that the student's program of study accesses the NC*SCS* and NCESS through the Extended Content Standards at the student's assigned grade level.
3. Determine that the student, who is in grade 3–8 or grade 10, will not participate in the statewide standard administration, with or without accommodations, of tests designated for the student's grade level. The IEP team must ensure that the decision for a student to participate in a statewide test administration or to participate in the ***NCEXTEND1*** is <u>not</u> the result of excessive or extended absences or social, cultural, or economic differences. These decisions (and the basis upon which they are made) must be documented in the student's IEP.
4. Address the consequences, if any, that participation in the ***NCEXTEND1*** may have on

the student's educational career, especially in meeting graduation requirements, if applicable.

5. Recommend student participation in the ***NCEXTEND1***.
6. Inform the parents/guardians that their child is being evaluated on the extended NC*SCS* by means of an alternate assessment with alternate academic achievement standards.

To determine participation in any of the ***NCEXTEND1*** alternate assessments, the following eligibility requirements must be met:

1. The student must have a current Individualized Education Program (IEP).
2. The student is enrolled in grades 3–8, 10, or 11 according to PowerSchool.
3. The student is instructed in the North Carolina Extended Standards in ALL assessed content areas.
4. The student has a SIGNIFICANT cognitive disability (i.e., exhibits 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).

The vast majority of students with disabilities do not have a significant cognitive disability. The ***NCEXTEND1*** is NOT appropriate for students who

- are being instructed in ANY OR ALL of the general grade-/course-level content standards of the North Carolina Standard Course of Study;
- demonstrate delays only in academic achievement;
- demonstrate delays owing primarily to behavioral issues;
- demonstrate delays only in selected areas of academic achievement; or
- if in high school, are pursuing a North Carolina high school diploma (including students enrolled in the Occupational Course of Study).

Note: Students who meet the ***NCEXTEND1*** eligibility requirements and are also identified as limited English Learners are exempt from the reading tests if they score below Level 4.0 Expanding on the reading subtest of the WIDA Access Placement Test (W-APTTM) and are in their first year in U.S. schools. These students, however, are required to participate in the administration of the math and science assessments.

In rare cases, a medical exception may be requested for medically fragile students who are unable to participate in the test administration because of a significant medical emergency and/or condition.

5.2 Assessor and Proctor Requirements

The ***NCEXTEND1*** alternate assessments are administered individually to each student by a trained Assessor with the presence of a proctor. Given the central role of the Assessors in ensuring test administration is conducted properly and scores are recorded correctly, the ***NCEXTEND1*** *Assessment Guide* clearly prescribes requirements for Assessors and proctors in order to ensure testing occurs fairly and uniformly. As stated in standard 6.1 of the *Standards,* "Test administrators should follow carefully the standardized procedures for administration and scoring specified by the test developer and any instructions from the test user" (p.114). In addition, Standard 4.15 of the *Standards* states

> *"The directions for test administration should be presented with sufficient clarity so that it is possible for others to replicate the administration conditions under which the data on reliability, validity, and (where appropriate) norms were obtained. Allowable variations in administration procedures should be clearly described. The process for reviewing requests for additional testing variations should also be documented"* (p. 90).

The *NCEXTEND1 Assessment Guide* requires that an Assessor is required:

- to be an employee of the school,
- to have training in the specific content area being assessed,
- to be familiar with the North Carolina Extended Content Standards, and
- to be the student's primary teacher for the assessed content area and to have routine contact with the student during classroom instruction.

The Proctor should:

- be familiar with the student being tested,
- be age 18 or older and not enrolled as a student in the K–12 public school system, and
- not be a parent or relative of the student being tested.

In addition to the specific requirements for Assessors and proctors, the Assessment Guide contains comprehensive details about every aspect of the assessment which are summarized in later sections.

5.3 *NCEXTEND1* Assessment Components

Each assessed academic discipline has a unique, grade-specific assessment packet. Each student has an assessment booklet that is used by the Assessor to present the test items and record scoring information. The assessment booklets contain all items for the content area at the student's grade level, the Assessor directions for item presentation, the script to be read with each item, the scoring criteria, and the Assessor Rating Sheet to record the observed student performance on the ***NCEXTEND1*** assessment items. The proctor also receives an identical assessment booklet for monitoring purpose. Each reading assessment at grades 3–8 and 10 utilizes a reading selection booklet that is unique to the specific grade level. The grade-specific reading selection booklets contain four selections.

The NCDPI also provides grade-level manipulative kits in each packet that are necessary to administer the ***NCEXTEND1*** assessment items. The Assessor receives one kit per grade level. Assessors whose students generally require adaptations to materials are allowed to preview manipulatives *in a secure setting* prior to the test administration to allow for time to make accommodations to materials. Additional manipulative kits are available upon request.

The items for the ***NCEXTEND1*** are performance-based items. Items are scored as correct with 2 score points for answering the item correctly the first time, 1 score point for answering the item correctly the second time, and 0 otherwise. The items for the ***NCEXTEND1***

are created to be as accessible as possible for all students. The number of items presented to each student is shown in *Table 3.3* to *Table 3.5* in Chapter 3.

5.4 Development and Review of Test Administration Procedures

The **NCEXTEND1** *Assessment Guide* is written to provide the information necessary for school administrators, test coordinators, and Assessors to implement a uniform administration. Included in the guide are:

- an overview of the relevant policies of the NC Statewide Testing Program;
- a description of the **NCEXTEND1** assessment;
- eligibility criteria for **NCEXTEND1** participation;
- the test administration process and instructions;
- responsibilities of Assessors; and
- information on the *Testing Code of Ethics*.

The ***NCEXTEND1*** *Assessment Guide* was reviewed internally by NCDPI staff and externally by Regional Accountability Staff.

5.5 *NCEXTEND1* Test Administration Training

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 described in the guides from the NCDPI Testing Policy and Operations Section once a year. Subsequently, the RACs provide training to LEA test coordinators on the processes for proper test administration. LEA test coordinators then provide training to school test coordinators. The training includes test security and testing procedures including 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 EL students), the ***NCEXTEND1*** Student Responses and Data Collection (NCEducation),

accommodated test administrations, test security (storing, inventorying, and returning test materials), and the *Testing Code of Ethics* (see Appendix 2-A).

5.6 Security Protocols Related to Test Administration

Test security is an ongoing concern in any testing program. When test security is compromised, it can undermine the validity of test scores. For this reason, the NCDPI has taken extensive steps to ensure the security of the ***NCEXTEND1*** alternate assessments by establishing protocols for school employees administering and handling paper tests.

5.6.1 Protocols for Assessor

Only school system employees are permitted as Assessors to administer secure state tests. Those employees must participate in the training for test administrators described in section 5.5. 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 read aloud all directions and information to students as indicated in the manual. The school test coordinator is responsible for monitoring test administrations within the building and responding to situations that may arise during test administrations.

5.6.2 Protocols for Handling and Administering Paper Tests

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

> *LEAs 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 specific items contained in, the test to persons other than authorize employees of the LEA.*

At the individual school, the principal is responsible for all test materials received. As established by SBE policy GCS-A-010, the *Testing Code of Ethics,* the principal must ensure test security within the school building and store the test materials in a secure, locked facility except when in use. The principal must establish a procedure to have test materials distributed immediately before each test administration. Every LEA and school must have a clearly defined system of check-out and check-in of test materials to ensure at each level of distribution and collection (LEA, school, and classroom) all secure materials are tracked and accounted for. LEA/charter school test coordinators must inventory test materials upon arrival from NCSU-TOPS and must inform NCSU-TOPS of any discrepancies in the shipment.

Before each test administration window, the building-level coordinator collects, counts, and stores all test materials in 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.

At the end of each test administration cycle, all testing materials must be returned to the school test coordinator according to directions specified in the *Assessment Guide*. Immediately after each test administration cycle, the school test coordinator shall collect, count, and return all test materials to the secure, locked facility. Any discrepancies must be reported immediately to the school system test coordinator. Upon notification, the school system test coordinator must report the discrepancies to the regional accountability coordinator and ensure all procedures in the Online Testing Irregularity Submission System are followed to document and report the testing irregularity. The procedures established by the school for tracking and accounting for test materials must be provided upon request to the school system test coordinator and/or the NCDPI Division of Accountability Services/North Carolina Testing Program.

At the end of the testing window, NCDPI mandates that all ***NCEXTEND1*** unused Assessor Booklets, unused Manipulative Card Kits, and unused Selection Booklets (intact teacher kits) be returned to NCDPI/NCSU-TOPS warehouse. Secure test materials are to be retained by the LEA in a secure (locked) facility with access controlled and limited to one or two

authorized school personnel only. After the required storage time (see *Table 5.1*) has elapsed, the LEA securely destroys these materials.

Table 5.1 Test Materials Designated to Be Stored by the LEA in a Secure Location

Test Material	Required Storage Time
All used answer sheets for operational tests	Six months after the return of students' test scores
NCEXTEND1 Used Assessor Booklets	Six months after the return of students' test scores
Original Braille writer/slate and stylus responses	Six months after the return of students' test scores
Original responses to a scribe	Six months after the return of students' test scores
Original responses using a typewriter or word processor	Six months after the return of students' test scores

5.6.3 Test Security Measures

Before test day, the Assessor must review each student's data with whom they have been matched as the Assessor, accessible via North Carolina online secured data platform referred to as NCEducation. The Assessor must contact the school test coordinator with questions related to data entry. After the test administration, the Assessor must enter the student's responses and complete the accommodations provided coding in NCEducation. The Assessor must be double checked by the designated school official to ensure error free data entry. The data entry must occur under secure conditions in a group setting. The Assessor must enter the student's responses while another individual verifies the data entry and a third individual acts as an objective observer of the process. All three individuals must sign the outside cover of the test book.

Student responses are securely sent when the Assessor clicks the submit button to the server at NC State University using the full HTTPS encryption process. Student records are transferred nightly to the NCDPI. These transfers are done following the NCDPI Secure File Transfer Protocol (SFTP) encryption rules and logic. More information on these processes can be

found in the NCDPI's *Maintaining the Confidentiality and Security of Testing and Accountability Data Guidance*. The NCDPI and NCEducation systems operate within the same network and are hosted at NC State University.

5.7 Administration

5.7.1 Assessment Window

All eligible students with significant cognitive disability are required to be administered the respective ***NCEXTEND1*** alternate assessments in ELA/reading and Mathematics at grades 3–8, Science at grades 5 and 8, and English II, Math I and Biology at Grade 10 within the final ten (10) instructional days of the school year. For 2014–15, LEAs/charter schools could apply for waivers and if granted, they could have five (5) additional instructional days. Exceptions was permitted to accommodate a student's IEP and Section 504 Plans.

5.7.2 Timing Guidelines

The *Standards* (2014) states *"although standardization has been a fundamental principle for assuring that all examinees have the same opportunity to demonstrate their standing on the construct that a test is intended to measure, sometimes flexibility is needed to provide essentially equivalent opportunities for some test takers"* (p. 51). In keeping with the *Standards* (2014), the NCDPI requires all alternate students be allowed ample opportunity to complete the assessments as long as they are engaged and working. The ***NCEXTEND1*** alternate assessments are administered individually to each student. The time required by a student to complete the assessment will be unique to each individual student, depending on the student's ability to maintain focus, his or her medical condition, and/or fatigue factor(s).

The ***NCEXTEND1*** alternate assessments may be administered over several days or may be completed in one session. The administration should align with the student's classroom experience. If a student routinely uses Multiple Testing Sessions during classroom instruction and similar classroom assessments, this accommodation should be documented in the student's IEP so appropriate planning and scheduling can take place before testing. The test design for the ***NCEXTEND1*** allows breaks to be taken at any time during testing as the need arises, regardless

of documentation in the student's IEP. The Assessor can use professional judgment to determine when a break is needed.

5.7.3 Testing Accommodations

State and federal law requires that all students, including students with disabilities (SWD) and students identified as EL, participate in the statewide testing program. Students may participate in the state assessments on grade level (i.e., general, alternate) with or without testing accommodations. Eligible students participating in the ***NCEXTEND1*** alternate assessments are provided with *"test accommodations if stated in their IEP to remove construct-irrelevant variance that otherwise would interfere with examinees' ability to demonstrate their standing on the target constructs."* (the *Standards*, 2014, p. 67) Testing accommodations are defined as *"changes in assessment materials or procedures that address aspects of students' disabilities that may interfere with the demonstration of their knowledge and skills on standardized tests"* (Thurlow & Bolt, 2001, p. 3). Accommodations are provided to eligible students together with appropriate administrative procedures to assure that individual student needs are met and, at the same time, maintain sufficient uniformity of the test administration.

For any state-mandated test, the accommodation for an eligible student must (1) be documented in the student's current IEP, Section 504 Plan, EL documentation, or transitory impairment documentation, and (2) the documentation must reflect routine use during instruction and similar classroom assessments that measure the same construct. When accommodations are provided in accordance with proper procedures as outlined by the state, results from these tests are deemed valid and fulfill the requirements for accountability.

According to Standard 6.2, "*When formal procedures have been established for requesting and receiving accommodations, test takers should be informed of these procedures in advance of testing"* (p. 115). In compliance with this, NCDPI specifies the following accommodations in North Carolina ***NCEXTEND1*** alternate assessments in the ***NCEXTEND1*** *Assessment Guide*:

- Braille Materials
- Large Print Materials created by the Assessor as needed for those students who routinely have print and visual materials enlarged.
- Assistive Technology Devices
- Interpreter/Transliterator Signs/Cues Test – For ELA grades 3–8 and 10, the Assessors reads selections 1–3 and all associated items to students. The use of Test Administrator Reads Test Aloud and /or Interpreter/Transliterator Signs/Cues Test accommodation for selection 4 will result in invalid scores.
- Magnification Devices
- Word-to-Word Bilingual (English/Native Language) Dictionary/Electronic Translator for students who are also identified as EL and have scored below 5.0 on the most recent ACCESS for EL or its alternate.
- Multiple Testing Sessions
- Adaptations to NCDPI-Provided Manipulatives such as raised lines, enlarged text/pictures, placement of pictures on information boards, and use of student-specific symbols. If a student requires and uses adapted materials routinely during instruction and this testing accommodation is documented in the student's IEP, the assessor may adapt the NCDPI-provided manipulatives as necessary before conducting the assessment. Assessors may access the manipulative cards under secure conditions in a group setting (i.e., three or more designated school personnel) up to two weeks before the test administration in order to make adaptations for those students who require this accommodation. Some examples of adapted materials are the use of assistive technology, large print cards, colored cards, and raised line cards. These types of materials should be used routinely in the classroom. Students requiring Braille cards should have Braille materials for the **NCEXTEND1** assessments ordered for them before the test administration.

The ***NCEXTEND1*** alternate assessments are to be read aloud to all students as specified in the Assessor booklet with the exception of ELA/reading selection 4. 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/. In addition, 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.

According to the *Standards* (2014), an appropriate accommodation addresses student's specific characteristics but does not change the construct the test is measuring or the meaning of scores. However, when necessary modifications that change the construct are provided to students to measure their standing on some intended construct, the modified assessment should be treated like a newly developed assessment. The NCDPI assessment guide recommends that students should only be allowed the same accommodations for assessments as those routinely used during classroom instruction and other classroom assessments that measure the same construct.

5.7.4 English Learners

Per State Board policy GCS-C-021, students identified as English Learners (EL)[9] must participate in the statewide testing program using the accommodated or non-accommodated standard test administration. The WIDA-ACCESS Placement Test (W-APT™) is the assessment used in North Carolina for initial identification and placement of students identified as EL. The assessment on the W-APT is based on results of the Home Language Survey (HLS) process (State Board policy GCS-K-000). The HLS process and the identification and subsequent placement of EL students in English as a Second Language (ESL) services are guided at the state level by the NCDPI Curriculum and Instruction Division. Additional information can be found at http://eldnces.ncdpi.wikispaces.net/Home+%28ELD%29.

[9] Once identified as EL based solely on the results of the W-APT™, the student is required by state and federal law to be assessed annually with the state-identified English language proficiency test. The test currently used by North Carolina for annual assessment of English Learners (ELs) is the Assessing Comprehension and Communication in English State-to-State for English Language Learners, or the ACCESS for ELLs®.

The Alternate ACCESS for ELLs is an option to the administration of the ACCESS for ELLs 2.0 test to students in grades 1–12 who are classified as ELs and have significant cognitive disabilities that prevent their meaningful participation in the ACCESS for ELLs 2.0 assessment. The Alternate ACCESS for ELLs is designed for only a small population of ELs who meet the following eligibility criteria:

- The student has a current IEP.
- The student participates in the general education curriculum through the Extended Content Standards.
- The student has a significant cognitive disability (i.e., exhibits 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).
- The student's ACCESS for ELs scores from the prior year yielded NA across any or all domains or yielded a composite score of less than 2.0. (If scores are 2.0 or above, the student does not qualify for the Alternate ACCESS for ELLs and must continue to take the regular ACCESS for ELLs.)
- If the student does not have ACCESS for ELLs testing for the prior year, the student's WIDA-ACCESS Placement test (W-APT) results must have a proficiency level of 1 in all applicable domains.

For ***NCEXTEND1*** assessments, EL students qualify to receive the following accommodations based on their scores on the WIDA-ACCESS Alternate Placement Test (W-APT™). The state approved EL testing accommodations for ELA include:

- Multiple testing session
- Student read aloud to self
- The state approved EL testing accommodations for math and science include:
- Multiple Testing Sessions
- Schedule Extended Time
- Student Reads aloud to self

- English/Native Language Word-to-Word Bilingual Dictionary/Electronic Translator

5.7.5 Student Participation

The Administrative Procedures Act 16 NCAC 6D. 0301 requires that all public school students enrolled in grades for which the SBE adopts an assessment, including every child with disabilities, participate in the testing program unless excluded from testing (16 NCAC 6G.0305(g)). For the ***NCEXTEND1***, the 1% of students identified by their IEP as having significant cognitive disability and who are taught based on the *North Carolina Extended Content Standards* are eligible for the ***NCEXTEND1*** alternate assessments in ELA, math, and science. The percentages of students who participated in the general assessment and the ***NCEXTEND1*** alternate assessment are presented in *Table 5.2* through *Table 5.4*. As required, ***NCEXTEND1*** students make up about 1% of the total students assessed at grades 3–8. For grade 10, the percentages are somewhat off because the parallel general assessments of English II, Math I, and Biology are not grade specific like their ***NCEXTEND1*** counterpart which are only administered to ***NCEXTEND1*** students enrolled in grade 10.

Table 5.2 ELA Percentage of Students by Assessment – General and ***NCEXTEND1***

Grade	2012–2013			2013–2014			2014–2015		
	General (%)	EXT1 (%)	All	General (%)	EXT1 (%)	All	General (%)	EXT1 (%)	All
3	99.05	0.95	104,037	99.05	0.95	112,245	98.96	1.04	117,596
4	99.06	0.94	111,197	99.03	0.97	104,571	99.03	0.97	115,075
5	99.03	0.97	110,781	99.00	1.00	112,296	98.95	1.05	107,720
6	98.99	1.01	112,710	99.02	0.98	112,058	98.96	1.04	115,663
7	99.04	0.96	111,854	98.98	1.02	114,182	99.01	0.99	115,810
8	98.94	1.06	110,024	98.91	1.09	113,181	98.87	1.13	118,081
10	99.22	0.78	106,613	99.18	0.82	110,473	99.20	0.80	115,609

*Table 5.3 Math Percentage of Students by Assessment – General and **NCEXTEND1***

Grade	2012–2013			2013–2014			2014–2015		
	General (%)	EXT1 (%)	All	General (%)	EXT1 (%)	All	General (%)	EXT1 (%)	All
3	99.05	0.95	104,583	99.06	0.94	113,080	98.96	1.04	117,624
4	99.06	0.94	112,037	99.03	0.97	104,995	99.03	0.97	115,086
5	99.03	0.97	111,677	99.01	0.99	112,838	98.95	1.05	107,738
6	99.00	1.00	113,391	99.02	0.98	112,572	98.96	1.04	115,674
7	99.05	0.95	112,402	98.98	1.02	114,584	99.01	0.99	115,809
8	98.94	1.06	110,369	98.91	1.09	113,478	98.87	1.13	118,069
10	99.29	0.71	117,823	99.22	0.78	117,373	99.22	0.78	119,731

*Table 5.4 Science Percentage of Students by Assessment – General and **NCEXTEND1***

Grade	2012–2013			2013–2014			2014–2015		
	Genera l (%)	EXT1 (%)	All	General (%)	EXT1 (%)	All	General (%)	EXT1 (%)	All
5	99.03	0.97	111,367	99.01	0.99	112,773	98.95	1.05	107,735
8	98.94	1.06	110,150	98.91	1.09	113,342	98.87	1.13	117,971
10	99.21	0.79	105,208	99.15	0.85	107,550	99.17	0.83	112,244

According to State Board policy GCS-A-001, school systems shall, at the beginning of the school year, provide information to students and parents or guardians advising them of the district-wide and state-mandated assessments that students are required to take during the school year. In addition, school systems must provide information to students and parents or guardians to advise them of the dates the tests will be administered and how the results from each assessment will be used. Information provided to parents about the tests must include whether the SBE or local board of education requires the test. School systems must report test scores and interpretative guidance from district-wide and/or state-mandated tests to students and parents or guardians within thirty (30) days of the generation of the score at the school system level or receipt of the score and interpretive documentation from the NCDPI.

5.7.6 Medical Exclusions

There may be rare circumstances in which a student with a significant medical emergency and/or condition may be excused from the required state tests. For requests that involve significant medical emergencies and/or conditions, the LEA superintendent or charter school director must submit a written request to the NCDPI. The request must include a detailed justification of why the student's medical emergency and/or conditions prevent participation in the respective test administration during the testing window and the subsequent makeup period. Most of what is submitted for the medical exception is housed at the school level (IEP, dates of the scheduled test administration[s] and makeup dates, number of days of instruction missed due to the emergency/condition, expected duration/recovery period, explanation of the condition and how it affects the student on a daily basis, etc.) The student's records remain confidential, and any written material containing identifiable student information is not disseminated or otherwise made available to the public. For more information on the process for requesting special exceptions based on significant medical emergencies and/or conditions, please review http://www.ncpublicschools.org/docs/accountability/1516medexcept.pdf.

5.8 Testing Irregularities

The test administrator or proctor must report any alleged testing violation or testing irregularity to the school test coordinator on the day of the occurrence. The school test coordinator must contact the school system coordinator immediately with any allegation of a testing violation. The school test coordinator must then conduct a thorough investigation and complete the Report of Testing Irregularity: Part 1, which is located in the back of the *Administrative Guide*. Part 1 of the irregularity must be completed and filed with the school system test coordinator within five days of the test administration. Different incidents must be documented on separate reports of testing irregularities. If the superintendent or school system test coordinator declares a misadministration, the school system must complete both sides of the Report of Testing Irregularity form prior to sending both sides of the form to the regional accountability coordinator (RAC). *All requested information on the form must be completed.*

Examples of testing irregularities include, but are not limited to, the following:

1. Failing to follow the procedures as described in the *Administrative Guide*;
2. Failing to follow the test schedule procedures designated by the NCDPI Division of Accountability Services/NC Statewide Testing Program;
3. Failing to test all eligible students (State Board of Education policy GCS-A-010);
4. Administering tests to ineligible students;
5. Interpreting, explaining, or paraphrasing the test directions or the test items [State Board of Education policy GCS-A-10 (16 NCAC 6D .0306)];
6. Giving students instruction related to the concepts measured by the tests on the morning of the test administration or during the test administration session;
7. Paraphrasing, omitting, revising, or rewriting the script or the directions contained within the test administration booklet;
8. Failing to return the originally distributed number of secure assessment materials (e.g., reading selection booklets, assessment booklets, *Administrative Guide*s) to designated school personnel;
9. Allowing school or district personnel who do not have a legitimate need access to the assessment;
10. Failing to administer the assessment documented on the IEP documentation;
11. Failing to follow timelines for assessment requirements; and
12. Removing materials (i.e., assessment booklets and manipulatives kits) from the designated location (i.e., school building).

5.9 Misadministration

School systems must monitor test administration procedures. According to State Board of Education policy GCS-A-001 (16 NCAC 6D .0302), if school officials discover any instance of improper administration and determine that the validity of the test results has been affected, they must notify the local board of education and order the affected students to be retested. If the school system discovers any instance of an improper administration and determines that the validity of the test results has been affected, a misadministration is declared. Only the

superintendent and the school system test coordinator have the authority to declare a misadministration at the local level. When a misadministration is declared, the affected students' scores must be deemed invalid. Any misadministration must be reported to the local board of education and the regional accountability coordinator (RAC). All decisions to invalidate scores on the basis of misadministration must be reported using the appropriate documentation.

5.10 Invalid Test Scores

In the event that procedures specified in this guide or in state accommodations publications are not followed during the actual test administration, the NCDPI Division of Accountability Services may declare the test scores as invalid. If the test scores are invalid, the results are not to be included in a student's permanent record, used for placement decisions, or used for student and school accountability.

Chapter 6 Scoring

The ***NCEXTEND1*** grade level assessments are designed and administered to a small and selective sample of students. As stated in chapter 5 (see *Table 5.2*) these assessments are administered to the 1% of students with significant cognitive disability at each grade level. Current Edition 3 of ***NCEXTEND1*** has one base form with 15 multiple-choice performance-based items for each subject at grades 3–8 and 10. The performance-based scoring rule adopted for these assessments, in combination with the smaller samples, does not support the use of large-scale psychometric methodology available in the Item Response Theory family. Also, the use of a single form excluded the need of equating and scaling.

This chapter describes the processes used for scoring ***NCEXTEND1*** items and the procedure adopted to create and report total scores. The first section of this chapter describes the scoring procedures for ***NCEXTEND1*** performance-based multiple-choice items. The second section describes the data certification processes used by NCDPI to ensure the quality of student data. The information in this Chapter is intended to comply with AERA, APA, & NCME (2014) Standard 4.18, which states:

> *Procedures for scoring and, if relevant, scoring criteria, should be presented by the test developer with sufficient detail and clarity to maximize the accuracy of scoring. Instructions for using rating scales or for deriving scores obtained by coding, scaling, or classifying constructed responses should be clear. This is especially critical for extended-response items such as performance tasks, portfolios, and essays"* (p. 91)

Information in this chapter is presented with enough detail to meet Standard 4.18, but not so much as to compromise the integrity of the test items.

6.1 Recording Student Responses

All student responses for ***NCEXTEND1*** items are recorded by test administrators during each administration session. As described in the ***NCEXTEND1****Assessment Guide*, each Assessor is provided with a test booklet on which they are to record student responses following

administration of each item. ***NCEXTEND1*** students are allowed up to two trials to provide the correct response for each item. If a student selects an incorrect response option during the first trial or the Assessor recorded a No Response, that response option, or another incorrect option as specified in the Assessor script, is pulled out from the foil list, and the item is presented with only two response options for a second trial.

Students may answer items by responding in ways used routinely in their daily instruction in the classroom (e.g., eye gazing, verbalizations, pointing, etc.). Using the Assessor booklet, Assessors must transcribe student responses for Trial 1 and Trial 2 (if necessary) for each item. The following response options are valid for Trial 1: "A," "B," "C," or "NR" (No Response). For Trial 2, the following response options are valid: "Not Used," "A," "B," "C," or "NR". If Trial 2 was not necessary, the default response of "Not Used" should be selected. When recording a student's response, Assessors are guided by the following rules:

- Trial 1 must have a response for every item.
- Trial 1 and Trial 2 cannot be marked as the same response (e.g., Trial 1: A, Trial 2: A)
- If Trial 1 is marked NR, Trial 2 must have a response.

After the test administration, the Assessor must enter the student's responses and complete the Accommodations Provided information in NC Education NCEXTEND1 data entry portal. The data entry must occur under a secure condition in a group setting with three or more designated school personnel. The Assessor is to enter the student's response, another individual is to verify the data entry, and a third individual is to act as an objective observer of the process. All three individual must sign the outside cover of the test book.

The NC Education data entry screen for each item also has a text box for Assessors to enter any comments documented in the margins of the Assessor booklet for that item. Comments may include remarks about item quality, clarity of language, alignment to the North Carolina Extended Content Standards, formatting and graphics quality, bias, and special accommodations made for a particular item.

Once all items have been recorded and all the quality checks verified including error messages generated by the system, the Assessor then clicks the submit button, and the student response data are saved and transferred to the secured server for processing.

6.2 Scoring Student Responses

The NCDPI WinScan software program is used for scoring ***NCEXTEND1*** responses. WinScan is a specialized scoring and reporting software program created and managed by the NCDPI Accountability Division. At the beginning of each testing window, a new release of WinScan is updated and distributed to all LEAs. Each version is programmed using the score keys and conversion tables for total scores to achievement levels.

Once the LEA test coordinator receives confirmation that all student data has been submitted through the NC Education Data Entry System, he/she runs the WinScan report to score and generate all the necessary school and student level reports. The scoring algorithm in WinScan matches each response to the answer key and awards a score point of "2" if the correct response was observed in Trial 1, a score of "1" if the correct response was recorded in Trial 2, a score of "0" for incorrect response, and missing for no response in either Trial 1 or 2.

Chapter 7 Operational Field-Test Data Analyses

This chapter describes the item and test analyses conducted after the 2012–13 operational field test administration of the ***NCEXTEND1*** assessments. The purpose of these analyses was to evaluate classical item statistics for the newly developed items prior to standard setting.

7.1 Operational Field-Test Sample 2012–13

In 2012–13 ***NCEXTEND1*** Edition 3 was administered as an operational field test in ELA and Math grades 3–8 and 10, and Science grades 5, 8, and 10 to students with significant cognitive disability. Demographic descriptive summaries for students who were administered ***NCEXTEND1*** during the 2012–13 operational field test are shown in *Table 7.1* for ELA, Table 7.2 for Math and *Table 7.3* for Science. On average about 1,050 students identified as having significant cognitive disability were administered an ***NCEXTEND1***assessment in grades 3–8 and 10. Across all grades there was a 2:1 ratio of males (about 67%) to females (about 33%). Overall a little over 7,372 students were administered ***NCEXTEND1*** in ELA and Math and about 3,104 students in Science.

Table 7.1 ELA Student Demographic for **NCEXTEND1** *Operational Field-Test 2012–13*

ELA Grade	N	Gender		Ethnicity						
		% Female	% Male	% Asian	% Black	% Hispanic	% American Indian	% Multiracial	% Native Hawaiian/Pacific Islander	% White
3	993	33.53	66.47	2.42	33.43	13.70	1.51	4.93	0.10	43.91
4	1,056	32.48	67.52	2.46	34.75	12.22	1.99	3.88		44.70
5	1,088	30.88	69.12	1.56	33.64	11.76	1.19	4.04		47.79
6	1,137	34.56	65.44	2.02	35.62	10.20	1.14	2.99	0.18	47.85
7	1,082	35.77	64.23	2.50	35.21	9.89	2.50	2.22	0.09	47.60
8	1,175	33.79	66.21	2.21	36.26	10.38	1.36	2.38		47.40
10	841	34.48	65.52	1.31	37.22	9.39	1.66	2.62	0.24	47.56
All	**7,372**	**33.63**	**66.37**	**2.09**	**35.13**	**11.08**	**1.61**	**3.28**	**0.08**	**46.72**

Table 7.2 Math Student Demographic for **NCEXTEND1** *Operational Field-Test 2012–13*

Math Grade	N	Gender		Ethnicity						
		% Female	% Male	% Asian	% Black	% Hispanic	% American Indian	% Multiracial	% Native Hawaiian/Pacific Islander	% White
3	993	33.53	66.47	2.42	33.43	13.70	1.51	4.93	0.10	43.91
4	1,057	32.45	67.55	2.46	34.72	12.20	1.89	3.88		44.84
5	1,087	30.91	69.09	1.56	33.67	11.78	1.20	4.05		47.75
6	1,139	34.59	65.41	2.02	35.56	10.36	1.14	2.99	0.18	47.76
7	1,084	35.70	64.30	2.58	35.24	9.87	2.49	2.21	0.09	47.51
8	1,176	33.76	66.24	2.21	36.22	10.46	1.36	2.38		47.36
10	842	34.44	65.56	1.31	37.17	9.38	1.66	2.61	0.24	47.62
All	**7,378**	**33.61**	**66.39**	**2.10**	**35.12**	**11.11**	**1.60**	**3.28**	**0.08**	**46.71**

Table 7.3 Science Student Demographic for **NCEXTEND1** *Operational Field-Test 2012–13*

Science Grade	N	Gender		Ethnicity						
		% Female	% Male	% Asian	% Black	% Hispanic	% American Indian	% Multiracial	% Native Hawaiian/ Pacific Islander	% White
5	1,087	30.91	69.09	1.56	33.58	11.78	1.20	4.05		47.84
8	1,175	33.79	66.21	2.21	36.17	10.47	1.36	2.38		47.40
10	842	34.44	65.56	1.31	37.17	9.38	1.66	2.61	0.24	47.62
All	**3,104**	**32.96**	**67.04**	**1.74**	**35.53**	**10.63**	**1.39**	**3.03**	**0.06**	**47.62**

7.2 Operational Field-Test Item Analyses

As noted in chapter 4, the 2012–13 ***NCEXTEND1*** forms were assembled using a combination of newly developed items and revised field test items from the 2011–12 stand-alone field test administration. After the operational field-test administration in 2013, NCDPI conducted classical item analyses to ensure items performed as expected and the overall statistical quality of the forms met NCDPI technical requirements.

Classical Test Theory (CTT) was chosen for statistical analyses of items and forms because the total number of students who were administered ***NCEXTEND1*** assessments in each grade was small (ranged from 841 to 1,175). For assessments administered to a smaller and selective sample, CTT statistics provides acceptable sample-based item summary estimates with which to evaluate the overall quality of items. Specifically, two CTT item level statistics were examined for each item:

- Pvalue: For all 15 performance based multiple-choice items on ***NCEXTEND1*** with valid score range of 0 – 2, Pvalue is defined as the average raw score divided by the maximum possible score for each item (2). For example, if the item average score is 1 point, the pvalue for the performance multiple-choice item is 0.50. (average observed score/maximum possible item points)
- Polyserial Correlation Coefficients: These are special cases of Pearson correlation coefficient describe the relationship between a quantitative variable and an ordinal or multistep variable. Polyserial coefficients provide evidence of how well each item on a test form correlates with the latent construct estimated by total score.

The performance-based scoring rule used allowed students a second trial if they did not respond correctly the first time. During the second trial, one incorrect response foil is removed and the item is presented with two foils. Thus, the probability of chance score increases from about .33 to .50 between trials. A correct response at trial one is worth 2 points and a correct response at trial two is worth 1 point. Table 7.4 through Table 7.6 show the average items answered correctly after first and second trial for each grade with expected pvalue. On average students were able to answer three additional items correctly during trial two. Expected pvalue between trials one and two increased by about 0.10.

Table 7.4 **NCEXTEND1** *ELA Average Items Answered Correctly at First and Second Trial.*

ELA	Average Items Answered Correct at 1st Attempt (2pts)	Average Additional Items Answered Correctly at 2nd Attempt (1pts)	Average Pvalue at 1st Attempt	Average Pvalue at 2nd Attempt
Grade 3	9	3	0.62	0.71
Grade 4	9	3	0.57	0.68
Grade 5	8	3	0.56	0.67
Grade 6	8	3	0.57	0.68
Grade 7	9	3	0.57	0.68
Grade 8	8	3	0.54	0.66
Grade 10	8	4	0.53	0.65

Table 7.5 **NCEXTEND1** *Math Average Items Answered Correctly at First and Second Trial.*

Math	Average Items Answered Correct at 1st Attempt (2pts)	Average Additional Items Answered Correctly at 2nd Attempt (1pts)	Average Pvalue at 1st Attempt	Average Pvalue at 2nd Attempt
Grade 3	7	4	0.49	0.62
Grade 4	8	4	0.51	0.64
Grade 5	7	4	0.47	0.61
Grade 6	7	4	0.50	0.63
Grade 7	7	4	0.48	0.62
Grade 8	7	4	0.44	0.58
Grade 10	7	4	0.45	0.59

Table 7.6 **NCEXTEND1** *Science Average Items Answered Correctly at First and Second Trial.*

Science	Average Items Answered Correct at 1st Attempt (2pts)	Average Additional Items Answered Correctly at 2nd Attempt (1pts)	Average Pvalue at 1st Attempt	Average Pvalue at 2nd Attempt
Grade 5	9	3	0.57	0.69
Grade 8	9	3	0.62	0.72
Grade 10	8	4	0.54	0.66

Summary descriptive CTT statistics from the 2012–13 operational field test are shown in *Table 7.7* through *Table 7.9* for ELA, math, and science respectively. These results show average pvalues by grade and content area from the operational field test was above 0.50. Only Grade 10 Math had an item with a pvalue of 0.34. For all other grades and subjects, observed pvalue ranged from 0.44 to 0.87.

Polyserial correlations from operational field-test indicates medium to high correlation between items and the construct measured in each assessment. All item total correlations are positive and range from 0.19 to 0.78.

Table 7.7 ELA **NCEXTEND1** *CTT Descriptive Statistics Operational Field Test 2012–2013*

ELA	N Items	Pvalue				Polyserial Correlation			
		Average	STD	Min	Max	Average	STD	Min	Max
Grade 3	15	0.71	0.04	0.65	0.79	0.70	0.06	0.54	0.76
Grade 4	15	0.68	0.06	0.59	0.82	0.60	0.11	0.29	0.74
Grade 5	15	0.67	0.07	0.51	0.76	0.63	0.09	0.42	0.77
Grade 6	15	0.68	0.08	0.52	0.81	0.59	0.10	0.39	0.73
Grade 7	15	0.68	0.07	0.51	0.77	0.65	0.13	0.35	0.76
Grade 8	15	0.66	0.08	0.48	0.78	0.58	0.12	0.32	0.72
Grade 10	15	0.65	0.04	0.57	0.71	0.60	0.08	0.37	0.73

Table 7.8 Math **NCEXTEND1** *CTT Descriptive Statistics Operational Field Test 2012–2013*

Math	N Items	Pvalue				Polyserial Correlation			
		Average	STD	Min	Max	Average	STD	Min	Max
Grade 3	15	0.62	0.07	0.51	0.72	0.56	0.13	0.19	0.68
Grade 4	15	0.64	0.06	0.51	0.72	0.55	0.08	0.43	0.68
Grade 5	15	0.61	0.06	0.49	0.69	0.56	0.07	0.45	0.67
Grade 6	15	0.63	0.07	0.52	0.74	0.51	0.15	0.23	0.68
Grade 7	15	0.62	0.08	0.44	0.71	0.56	0.11	0.32	0.71
Grade 8	15	0.58	0.09	0.45	0.70	0.48	0.13	0.22	0.68
Grade 10	15	0.59	0.11	0.34	0.74	0.53	0.14	0.30	0.73

Table 7.9 Science **NCEXTEND1** *CTT Descriptive Statistics Operational Field Test 2012–2013*

Science	N Items	Pvalue				Polyserial Correlation			
		Average	STD	Min	Max	Average	STD	Min	Max
Grade 5	15	0.69	0.09	0.49	0.82	0.62	0.15	0.25	0.78
Grade 8	15	0.72	0.10	0.52	0.87	0.66	0.10	0.46	0.78
Grade 10	15	0.66	0.10	0.50	0.84	0.58	0.12	0.26	0.71

7.3 Test Reliability

Finally, reliability estimates were computed for all operational field test ***NCEXTEND1*** forms. The purpose was to ensure students' reported scores were dependable and that a significant proportion of the differences in performance among students could be explained by actual differences in the constructs assessed as opposed to random measurement error. In more general terms, reliability of test scores refers to the consistency or probability that if a student was given multiple administration of the same assessment, they will obtain about the same score each time. Reliability is usually reported either in terms of Standard Error of Measurement

(SEM) or measure of internal consistency coefficient based on Cronbach's alpha (Cronbach, 1951). Cronbach's alpha is calculated as:

$$\hat{\alpha} = \frac{\kappa}{\kappa - 1}(1 - \frac{\Sigma\hat{\sigma}_i^2}{\hat{\sigma}_X^2})$$

Where

-k is the number of items on the test form,

- $\hat{\sigma}_i^2$ is the variance of item i, and

- $\hat{\sigma}_X^2$ is the total test variance.

In CTT, SEM is sample based and it is estimated to be the same across the entire score range. Table 7.10 displays the Cronbach alpha reliability estimates with associated SEM for all *NCEXTEND1* operational field test forms. Reliabilities for ELA ranged from 0.76 to 0.89 with an average SEM of about 2.7 points on a 30-point scale. Reliabilities for Math were slightly lower with an average of around 0.72, particularly for grades 6 and 8 where the estimated reliabilities were less than 0.70. The average SEM was about 2.89 score point. Reliabilities across the three Science grades were generally acceptable. All but grade 10 Science (0.76) recorded reliability estimates that were above 0.80. Reliability estimates computed from 2013–14 and 2014–15 samples are consistent with these reported for 2012–2013 operational field-test.

In general, reliability estimates for ***NCEXTEND1*** assessments are noticeably below the NCDPI acceptable threshold of 0.85. Some leading contributing factors to the lower reliability estimate include: shorter test (15 items), heterogeneous makeup of student population due to varying complexities of their cognitive disabilities, variation in administration and Assessors.

Table 7.10 Reliability and SEM for **NCEXTEND1** *Operational Field Test 2012–2013*

Subject	Grade	Number of Items	MPS	2012–2013		2013–2014		2014–2015	
				Cronbach Alpha	SEM	Cronbach Alpha	SEM	Cronbach Alpha	SEM
ELA	3	15	30	0.89	2.47	0.88	2.43	0.87	2.45
	4	15	30	0.81	2.7	0.81	2.71	0.80	2.69
	5	15	30	0.83	2.71	0.81	2.69	0.82	2.69
	6	15	30	0.79	2.74	0.81	2.70	0.80	2.67
	7	15	30	0.85	2.60	0.82	2.63	0.83	2.61
	8	15	30	0.78	2.79	0.76	2.78	0.77	2.73
	10	15	30	0.81	2.78	0.81	2.75	0.83	2.72
Math	3	15	30	0.76	2.85	0.75	2.83	0.71	2.86
	4	15	30	0.74	2.91	0.77	2.86	0.74	2.89
	5	15	30	0.75	2.84	0.70	2.87	0.72	2.85
	6	15	30	0.69	2.93	0.71	2.89	0.69	2.88
	7	15	30	0.74	2.87	0.71	2.87	0.72	2.84
	8	15	30	0.64	2.98	0.60	2.99	0.63	2.93
	10	15	30	0.70	2.90	0.67	2.89	0.73	2.89
Science	5	15	30	0.81	2.65	0.79	2.63	0.81	2.63
	8	15	30	0.84	2.51	0.83	2.48	0.82	2.47
	10	15	30	0.76	2.78	0.76	2.74	0.80	2.69

Note: MPS – Maximum possible score

Chapter 8 Standard Setting

Standard setting is a process used to set achievement (proficiency) levels. Standard setting is recommended whenever an assessment system undergoes major revisions or changes to the underlying standards, as was the case in 2010 with the adoption of the new extended NC*SCS* and NC*ESS* and the development of the READY accountability assessment system to measure students' college- and- career readiness. In July 2013, after the first operational field test administration of the ***NCEXTEND1*** assessments, the NCDPI contracted with Alpine Testing Solution, Inc. (Alpine) to conduct a standard-setting workshop. During the workshop, educators recommended cut scores and achievement levels for ***NCEXTEND1*** Edition 3 assessments.

8.1 *NCEXTEND1* Standard Setting Overview

Standard 5.21 (AERA, APA, & NCME, 2014) states that *"when proposed score interpretation involves one or more cut scores, the rationale and procedures used for establishing cut score should be documented."* Standard setting is a process used to define achievement levels and the cut scores corresponding to those levels with associated achievement level descriptors (Extended Content Standards). A cut score is simply the score students must meet in order to be in a particular achievement level. The cut score separates students whose score is below the cut score into one level and those whose scores are at or above the cut score into the next and higher level.

Between July 30 and August 1, 2013, after the operational field-test administration of the ***NCEXTEND1***, a total of 44 North Carolina ELA, math, and science educators (15 for elementary, 15 for middle school, and 14 for high school) convened in Chapel Hill, North Carolina, to make cut score recommendations for the ***NCEXTEND1*** Edition 3 assessments. In addition, a subset of 12 elementary, middle, and high school panelists were asked to examine the vertical continuity of the impact data (percentage of students in each achievement level) associated with the recommended cut scores.

The Extended Angoff method (Plake & Hambleton, 2001) was used by panelists in a series of rounds to recommend cut scores. Alpine staff designed and led the workshop. The full report of the standard setting can be found in Appendix 11-A.

At the conclusion of the standard setting workshop, three recommended cut scores with four achievement levels were present to the SBE for adoption. An abbreviated version of the final standard setting study prepared by Alpine[10] for the NCDPI is presented in this chapter.

8.1.1 Workshop Panelists and Background

Prior to the workshop, NCDPI provided information about eligible panelists who were then recruited by Alpine to participate in each grade span panel. Each grade span panel included 14–15 content experts from across the state (Jaeger, 1991; Raymond & Reid, 2001). Each panel included teachers who had experience with the Extended Content Standards, teachers who had experience working with students with disabilities, and general education teachers across subject areas.

The panelists' experience as educators with the Extended Content Standards as well as their gender distributions are summarized in *Table 8.1*. As illustrated by the table, a significant proportion of educators have experience with the Extended Curriculum. The panel also consisted of educators experienced with students with disability and general education. The panel consisted of a higher proportion of female educators.

Table 8.1 Panelist Experience as Educators

Panel	Experience			Gender		Total
	Experience with Extended Curriculum	Experience with SwD (not EC)	General Education Teacher	Male	Female	
Elementary School	12	1	2	3	12	15
Middle School	13	0	2	2	13	15
High School	3	8	3	3	11	14
Vertical Moderation	10	0	2	4	8	12

[10] Copyright © 2013, Alpine Testing Solution, Las Vegas NV

The panelists' professional degrees are summarized in *Table 8.2*. Panelists in elementary school and vertical articulation had bachelors' and masters' degrees, and the middle and high school panelists had bachelors', masters' and doctorates degrees.

Table 8.2 Panelist Professional Background: Three-GradePanels

Panel	Number of Panelists	Degree			Mean Years of Experience
		Bachelors	Masters	Doctorate	
Elementary School	15	6	9	0	13.3
Middle School	15	6	8	1	14.3
High School	14	5	8	1	16.7
Vertical Moderation	12	4	8	0	12.75

8.1.2 Vertical Articulation Committee

A subset of the elementary and middle school panelists was selected to participate in a facilitated discussion of the vertical continuity of the impact data associated with the recommended cut scores. Specific gender, experience, and professional background information of this subgroup is also provided in *Table 8.1* and *Table 8.2*.

8.1.3 Workshop Orientation

On the first day of the workshop, a general orientation was held for all panelists. NCDPI staff welcomed the group. An Alpine Psychometrician provided an orientation that covered the purpose of the workshop, the goals of the workshop, and the processes that would be used to accomplish each goal. Following the orientation, panelists worked within smaller grade span panels for the remainder of the workshop.

8.2 Achievement Level Descriptors

To begin creating the achievement level descriptors (ALDs), panelists were divided into table groups with representation from the diversity of the participants. Each group was assigned one or two sets of ALDs to draft based on general policy level descriptors, an example provided by NCDPI, and an example presented from another state's ALDs. In addition, the panelists were

told that their ALDs should focus at the transition point or threshold from one achievement level to the next (as opposed to policy, range, or reporting ALDs). This focus was to help panelists begin to think about how students perform at the transition points between adjacent levels of achievement. Within their respective subgroup, they listed ideas for each achievement level of the types of things a student at that level could do related to the Extended Content Standards for that grade level and subject area. The draft ALDs were then transferred to an electronic format so they could be shared with each grade level panel. Within each panel, the ALDs were reviewed for clarity and continuity across grade levels and subject areas.

As part of the ALD development process, a vertical articulation process was also included. Specifically, this included members of the elementary school grade-span panel meeting with members of the middle school grade-span panel to discuss the transition from grades 5 to 6 for ELA and Mathematics. Similar discussions were held with the middle school grade-span panel and the high school grade-span panel for ELA, mathematics, and science to ensure continuity and increasing expectations across the grade levels. Feedback from these cross-panel discussions was then shared with the original grade-span groups to inform any additional revisions to the ALDs.

8.3 *NCEXTEND1* Standard Setting

"Just Barely" Level Descriptors

The recommended range of cut scores is based on the Extended Angoff method (Plake & Hambleton, 2001). In this process, panelists are presented with the assessment just as students would see it and are asked to make item-level judgments. For each item, they are asked to imagine the "target student" and make their best judgment as to what score the student would likely achieve on each item (0 points, 1 point, 2 points). In this application, there were three groups of target students: students that are barely level 2, students that are barely level 3 and those that are barely level 4. By focusing on the transition points between the achievement levels (e.g., barely level 3 differentiates between levels 2 and 3), panelists imagine their expectations

for students who represent the minimum level of knowledge and skills at each of the achievement levels.

Panelists then articulate their expectations for the target student by rating each item. For each item, panelists predict the number of points that a target student will obtain on each item. Facilitators collect ratings and compute the panel-level statistics. The facilitator shared group median cut scores, the range of cut scores across the panel (including a graphical representation of the distribution), the estimated impact if the median cut scores were used (i.e., what percent of students would be classified at or above each achievement level), and the average item score from the spring 2013 administration year.

In Round 2, the group discussed two items for each assessment – one that was generally easier for students and one that was more difficult–to help with understanding of how to apply the ALDs to the rating task. After explaining this feedback, the facilitator instructed the panelists to review their first round of ratings and make any modifications they felt necessary in their second round of ratings. The second-round ratings were used to compute the final recommended cut scores. Once ratings were completed for all assessments, the final activity for the full group of panelists was the completion of an evaluation form designed to measure the level of confidence in the standard setting activities and their cut score recommendations. After the evaluations were completed, each participant was provided with a certificate of participation and the respective workshop was concluded.

8.4 Vertical Articulation Discussion

During the afternoon of the last day, a subset of 12 panelists from the elementary, middle, and high school panels were convened to discuss the continuity across grade levels within a subject area. This discussion included ELA, Math, and Science as separate topics. After showing panelists the impact results from the second round of ratings, the panel discussed a number of questions regarding interpretation and explanation of the results. Some of the questions posed to the group during this discussion included whether the impact across grade levels for a given subject area appeared reasonable. In addition, panelists were asked whether any grade levels appeared unreasonably high or low in terms of expectations. Some of the context included in the

discussion was the alignment of the ELA and Math assessments to the Extended Content Standards of the *Common Core State Standards*.

In general, panelists provided feedback suggesting that content expectations from elementary to middle school and eventually high school increased at a trajectory that is steeper than the typical progression of development for students who take the ***NCEXTEND1*** assessments. Further, there is a shift in cognitive complexity from more concrete to more abstract concepts in moving from elementary to middle school, particularly grades 6 to 7 in mathematics. There were some comments regarding the performance of students in the elementary grade levels in ELA being potentially higher than expected given the change in the expectations for students in the Extended Content Standards. Another point raised by panelists in the discussion was the influence of guessing on student performance. Given the design of the assessment administration, students had a reasonable probability of earning points on a given item through chance. There was consensus, almost unanimity, among the panelists that students would guess on items. This additional factor led NC DPI to consider including a guessing adjustment in the final recommendations to ensure that scores correspond with the meaning of the achievement levels.

8.5 Standard-Setting Results

The standard setting included two rounds of judgments. The results for each grade level are presented in Table 8.3 to

Table 8.5 for English Language Arts, Mathematics, and Science assessments, respectively. From the first round of ratings, each table includes the median recommended cut score (R1-Median) for each level along with the estimated impact (R1-Impact, percent of students at or above each performance level). From the second round of ratings, each table includes the median recommended cut score (R2-Median) for each level along with the estimated impact (R2-Impact, percent of students at or above each performance level), the standard deviation of the recommended cut scores (R2-SD) which represents the variability among the panel, and the range of recommended cut scores (R2-Range), which was estimated using the variability among the panel. Specifically, the range of recommended cut scores is estimated as:

High End of the Range = Median + 2 Standard Error of the Median Low End of the Range = Median - 2 Standard Error of the Median where,

$$SE_{Median} = \frac{\sigma}{\sqrt{N}}$$

The full results are shown graphically in Appendix B of the ***NCEXTEND1*** Standard Setting Technical Report.

Table 8.3 ELA **NCEXTEND1** *Standard-Setting Results by Grade and Performance Level*

Level	Result (Round/Grade)	3	4	5	6	7	8	10
	R1-Median	10	8	7	11	11	12	6.5
	R1-Impact	94%	97%	96%	95%	94%	92%	97%
2	R2-Median	10	8	7	11	11	11	8.5
	R2-Impact	94%	97%	96%	95%	94%	94%	96%
	R2-SD	3.20	3.02	2.15	3.71	3.72	3.94	5.14
	R2-Range	8-12	6-10	6-8	9-13	9-13	8-14	5-12
	R1-Median	22	20	18	21	22	21	16
	R1-Impact	52%	60%	66%	52%	48%	48%	75%
3	R2-Median	20	20	17	20	22	20	18
	R2-Impact	60%	60%	72%	58%	48%	53%	61%
	R2-SD	3.22	3.58	2.6	2.47	2.31	2.23	4.45
	R2-Range	18-22	18-22	15-19	18-22	21-23	19-21	15-21

	R1-Median	27	26	26	27	25	26	23.5
	R1-Impact	34%	26%	24%	16%	34%	18%	30%
4	R2-Median	28	26	25	27	25	26	25
	R2-Impact	29%	26%	29%	16%	34%	18%	26%
	R2-SD	1.47	2.32	1.98	2.19	1.79	2.02	2.18
	R2-Range	27-29	25-27	24-26	26-28	24-26	25-27	24-26

Table 8.4 Mathematics **NCEXTEND1** *Standard-Setting Results by Grade and Performance Level*

Level	Result (Round/Grade)	3	4	5	6	7	8	10
	R1-Median	7	7	6	8	12	12	5
	R1-Impact	96%	97%	96%	97%	93%	92%	97%
2	R2-Median	7	7	6	9	12	10	6
	R2-Impact	96%	97%	96%	96%	93%	95%	96%
	R2-SD	2.64	1.64	2.02	4.22	3.06	2.08	2.95
	R2-Range	5-9	6-8	5-7	6-12	10-14	9-11	4-8
	R1-Median	20	19	19	22	22	22	15
	R1-Impact	46%	52%	50%	32%	29%	19%	77%
3	R2-Median	20	18	19	20	23	21	15
	R2-Impact	46%	59%	50%	47%	22%	26%	77%
	R2-SD	3.87	2.72	3.17	4.22	1.77	2.2	3.3
	R2-Range	18-22	16-20	17-21	17-23	22-24	20-22	13-17
	R1-Median	26	26	25	27	27	27	22
	R1-Impact	13%	13%	15%	5%	6%	2%	21%
4	R2-Median	27	26	25	27	28	27	23
	R2-Impact	7%	13%	15%	5%	4%	2%	15%
	R2-SD	3.10	2.08	2.40	2.54	1.61	2.33	2.49
	R2-Range	25-29	25-27	23-27	25-29	27-29	25-29	21-25

Table 8.5 Science **NCEXTEND1** *Standard-Setting Results by Grade and Performance Level*

Level	Result (Round/Grade)	5	8	10
2	R1-Median	8	10	9
	R1-Impact	96%	96%	96%
	R2-Median	9	11	9
	R2-Impact	96%	96%	96%
	R2-SD	1.87	2.90	4.53
	R2-Range	8-10	9-13	6-12
3	R1-Median	21	21	19
	R1-Impact	54%	58%	62%
	R2-Median	21	22	19
	R2-Impact	54%	54%	62%
	R2-SD	1.87	2.50	4.60
	R2-Range	20-22	20-24	16-22
4	R1-Median	27	27	25
	R1-Impact	18%	28%	20%
	R2-Median	25	27	25
	R2-Impact	30%	28%	20%
	R2-SD	1.66	1.78	2.56
	R2-Range	24-26	26-28	23-27

8.6 Standard-Setting Guessing Adjustment

Given the nature of the administration and scoring of the ***NCEXTEND1*** assessments (e.g., three choices to select from, followed by a second chance with only two choices), there is a reasonable probability of students earning some points on this exam by simply guessing. Because the standard-setting panelists were instructed to estimate how the students would perform on the items using their knowledge, skills, and abilities, without guessing, the suggested adjustment applied is based on the probability of a student earning points on those items that they would answer incorrectly due to lack of knowledge, skills, or abilities. A full description of the guessing adjustment can be found in the full report. This guessing adjustment was applied consistently across grade levels and subject areas with one notable exception, Grade 10 Mathematics. In reviewing the median recommended results from the high school panel, panelists observed that for the Level 3 cut score (i.e., students are meeting the standard), the panel's median recommendation was at the chance level– 15 points of a possible 30. After reviewing recommendations across grade levels and subject areas,

panelists noted that this was the only recommendation that occurred at the chance or lower level. As a result, Alpine recommended a two-phase guessing adjustment for the Grade 10 Mathematics Level 3 cut score.

Specifically, as a first phase of the adjustment, panelists recommended raising the group's recommendation to chance plus one score point, which resulted in a median recommendation of 16 as opposed to the group's initial recommendation of 15. Given the standard error of the median associated with the group's recommendations, this increase falls within the 95% confidence interval for what we might expect. The second phase was to then apply the guessing adjustment described above that was applied across grade levels and subject areas. This additional step for the Grade 10 Mathematics assessment at Level 3 was intended to be consistent with expectations for meeting the standard across grade levels and also apply the same statistical adjustment. The median and adjusted median cut scores are shown in *Table 8.6* through *Table 8.8* and impacts are shown in *Figure 8.1* to *Figure 8.3*. As can be seen from the tables and figures, the median and adjusting median were significant and their impacts were also significant, more so for Level 2 followed by Level 3. Level 4 has a 1-score point adjustment for most grades and contents.

Table 8.6 ELA Recommended Cut Scores and Impact Adjusted for Guessing

Level	Result	Grade						
		3	4	5	6	7	8	10
2	Median	10	8	7	11	11	11	9
	Median-Adj	18	17	17	18	18	18	17
	Impact	93.80%	96.89%	95.70%	94.93%	94.22%	93.65%	95.86%
	Impact-Adj	68.70%	73.40%	72.28%	68.53%	66.70%	65.28%	68.56%
3	Median	20	20	17	20	22	20	18
	Median-Adj	23	23	20	23	24	23	21
	Impact	59.50%	59.53%	72.28%	58.04%	47.61%	53.01%	61.11%
	Impact-Adj	49.30%	41.51%	53.80%	39.95%	37.71%	37.34%	43.97%
4	Median	28	26	25	27	25	26	25
	Median-Adj	29	27	26	28	26	27	26
	Impact	29.30%	26.04%	29.37%	16.26%	33.76%	17.87%	26.24%
	Impact-Adj	21.80%	19.62%	24.43%	10.84%	28.81%	12.70%	21.87%

Table 8.7 Mathematics **NCEXTEND1** *Recommended Cut Scores and Impact Adjusted for Guessing*

Level	Result	Grade						
		3	4	5	6	7	8	10
2	Median	7	7	6	9	12	10	6
	Median-Adj	17	17	17	17	18	18	17
	Impact	96.00%	97.27%	96.25%	96.42%	93.03%	95.01%	96.46%
	Impact-Adj	67.60%	69.27%	64.56%	70.33%	59.95%	49.75%	59.62%
3	Median	20	18	19	20	23	21	16
	Median-Adj	23	21	22	23	25	23	20
	Impact	45.50%	58.91%	49.73%	47.03%	22.09%	25.97%	68.00%
	Impact-Adj	27.20%	38.36%	28.48%	24.52%	13.57%	14.13%	34.59%
4	Median	27	26	25	27	28	27	23
	Median-Adj	28	27	26	28	29	28	25
	Impact	7.00%	13.48%	14.84%	5.15%	3.85%	2.28%	14.99%
	Impact-Adj	4.00%	11.12%	10.07%	2.62%	2.11%	1.44%	8.62%

Table 8.8 Science **NCEXTEND1** *Recommended Cut Scores and Impact Adjusted for Guessing*

Level	Result	Grade		
		5	8	10
2	Median	9	11	9
	Median-Adj	17	18	17
	Impact	96.06%	95.60%	96.45%
	Impact-Adj	75.09%	75.21%	74.32%
3	Median	21	22	19
	Median-Adj	23	24	22
	Impact	54.49%	53.98%	61.66%
	Impact-Adj	42.22%	45.18%	40.24%
4	Median	25	27	25
	Median-Adj	26	28	26
	Impact	30.13%	27.66%	20.47%
	Impact-Adj	24.63%	19.80%	15.86%

Figure 8.1 Impact of ELA **NCEXTEND1** *Recommended Cut Scores and Guessing Adjustment (Adj)*

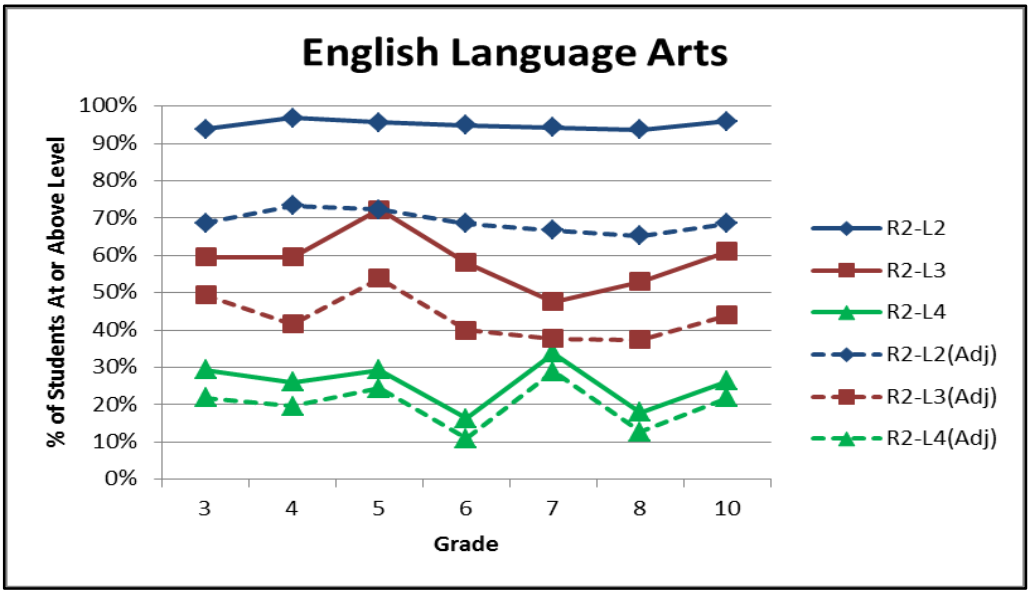


Figure 8.2 Impact of Mathematics **NCEXTEND1** *Recommended Cut Scores and Guessing Adjustment (Adj)*

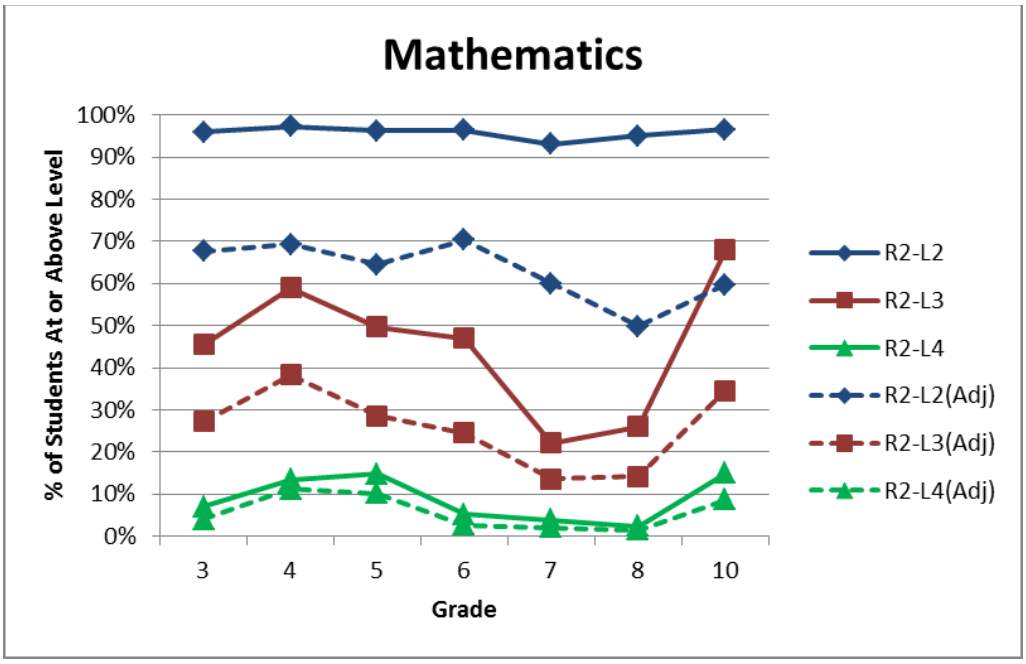
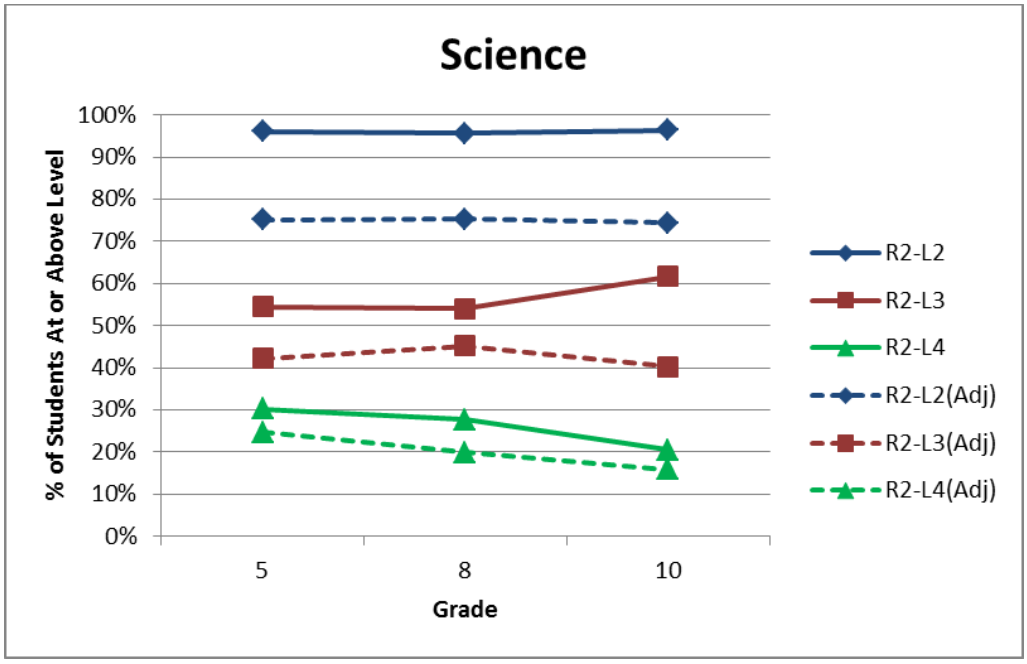


Figure 8.3 Impact of Science **NCEXTEND1** *Recommended Cut Scores and Guessing Adjustment* (Adj) *(Adj)*



8.7 Evaluation of Standard Setting

Each panelist responded to a series of evaluation questions about the various components of the workshop. The median response for each panel for each evaluation question is shown in *Table 8.9.* The overall results suggest that each panel felt the workshop was very successful in arriving at appropriate recommended cut scores. In addition to the closed-ended questions, panelists were allowed to provide comments about the workshop. These comments are included in Appendix E of the ***NCEXTEND1*** Standard Setting Technical Report.

Table 8.9 Median Evaluation Responses

Grade-Level Panel	Elementary	Middle	High School
Successfulness of training [6 = Very Successful to 1 = Very Unsuccessful]			
1a. Successfulness of orientation	5	5	5
1b. Successfulness of training on Yes/No method	5	5	4.5
1c. Successfulness of description of target students	5	5	5
1d. Successfulness of practice with method	6	5	5
1e. Successfulness of interpretation of feedback	5	5	5
1f. Successfulness of overall training	5	5	5
Time allocated to training [6 = Totally Adequate to 1 = Totally Inadequate]			
2a. Time – orientation	6	6	5
2b. Time – training on Yes/No method	6	4	5
2c. Time – description of target students	6	5	5
2d. Time – practice with method	6	5	5
2e. Time – interpretation of feedback	6	5	5
2f. Time – Overall training	6	5	5
Round One Yes/No Judgments			
3. Confidence in predictions [4 = Confident to 1 = Not at all confident]	4	3	3
4. Time for predictions [4 = More than enough time to 1 = More time needed]	4	3	3
Round Two Yes/No Judgments			
5. Confidence in predictions [4 = Confident to 1 = Not at all Confident]	4	4	4
6. Time for predictions [4 = More than enough time to 1 = More time needed]	4	3	3
Overall workshop			
7. Confidence in cut scores [4 = Confident to 1 = Not at all Confident]	3.5	4	3
8. Most useful feedback data *(mode reported)* [4 = Panel summary, 3 = Group discussions, 2 = Impact, 1=P-values]	3	2	1
9. Least useful feedback data *(mode reported)* [4 = Panel summary, 3 = Group discussions, 2 = Impact 1,=P-values]	4	4	3
10. Overall success [4 = Very Successful to 1 = Very Unsuccessful]	4	4	3
11. Overall organization [4 = Very Organized to 1 = Very Unorganized]	4	4	3

8.8 Validity of the Standard Setting

It is important to highlight the critical elements that provide validity evidence for the results of this standard setting. Kane's (1994, 2001) framework for standard setting validity evidence identifies three elements of validity evidence for standard settings: procedural, internal, and external. Procedural validity evidence for these studies can be documented through the careful selection of representative, qualified panelists, use of a published standard-setting method, completing the study in a systematic fashion, and collecting evaluation data that indicates the panelists felt they were confident in the cut score recommendations they made. Internal validity evidence suggested that panelists had similar expectations for the performance of the target students. This type of evidence is provided by the reasonable standard errors in the recommended cut scores for the second round of the standard-setting process. The final type of validity evidence, external, can be provided by triangulation with results from some other estimation of appropriate cut scores from outside the current standard-setting process and consideration of other factors that can influence the final policy.

8.9 Academic Achievement Standards Adoption and Revision

In October 2013, the SBE adopted College-and-Career Readiness Academic Achievement Standards and Academic Achievement descriptors for the End-of-Grade (EOG) and End-of-Course (EOC) regular and ***NCEXTEND1*** alternate assessments. After considering much input on the importance of having more for student achievement in the reported levels, the SBE adopted, at its March 2014 meeting, a methodology to add a new achievement level. With this additional achievement level, beginning in 2013–14 student performance on the ***NCEXTEND1*** assessments were reported based on five achievement levels as described in *Table 8.10* and *Table 8.11*.

Table 8.10 Revised 5 Achievement Levels Descriptors

Revised Achievement Level	Meets On-Grade-Level Proficiency Standard	Meets College-and Career-Readiness Standard
Level 5 denotes **Superior Command** of knowledge and skills.	Yes	Yes
Level 4 denotes **Solid Command** of knowledge and skills.	Yes	Yes
Level 3 denotes **Sufficient Command** of knowledge and skills.	Yes	No
Level 2 denotes **Partial Command** of knowledge and skills.	No	No
Level 1 denotes **Limited Command** of knowledge and skills.	No	No

Table 8.11 **NCEXTEND1** *Raw Score Cuts Based on Five Achievement Levels 2014 and Beyond*

Subject	Grade	Level 1	Level 2	Level 3	Level 4	Level 5
NCEXTEND1 *(ELA)*	3	≤15	16-20	21-22	23-28	≥29
	4	≤13	14-19	20-22	23-26	≥27
	5	≤13	14-16	17-19	20-25	≥26
	6	≤14	15-19	20-22	23-27	≥28
	7	≤14	15-20	21-23	24-25	≥26
	8	≤14	15-19	20-22	23-26	≥27
NCEXTEND1 *English II*	10	≤13	14-17	18-20	21-25	≥26
NCEXTEND1 *Mathematics*	3	≤13	14-19	20-22	23-27	≥28
	4	≤13	14-17	18-20	21-26	≥27
	5	≤13	14-18	19-21	22-25	≥26
	6	≤13	14-19	20-22	23-27	≥28
	7	≤14	15-21	22-24	25-28	≥29
	8	≤14	15-19	20-22	23-27	≥28
NCEXTEND1 *Math I*	10	≤13	14-16	17-19	20-24	≥25
NCEXTEND1 *Science*	5	≤13	14-19	20-22	23-25	≥26
	8	≤14	15-20	21-23	24-27	≥28
NCEXTEND1 *Biology*	10	≤13	14-18	19-21	22-25	≥26

The old level 4 became the new level 5 “Superior Command,” and students who scored at this level are considered to have met the on-grade-level proficiency standards and also prepared to be successful at the next level as defined by the Extended Content Standards and ALDs. The old level 3 became the new level 4 “Solid Command,” and students who scored at this level are also considered to have met the on-grade-level proficiency standards and also prepared to be successful at the next level as defined by the Extended Content Standards and ALDs.

The new Achievement Level 3 “Sufficient Command” identifies students who met on-grade–level-proficiency standard but are not prepared to be successful at the next level as defined by the Extended Content Standards and ALDs. This distinction assists schools in the delivery of differentiated instruction that best meets the needs of all student. The new Level 3 minimum scale score was created by subtracting one standard error of measurement from the original Level 3 scale score. Level 1 “Limited Command” and Level 2 “Partial Command” remained unchanged and describes students who have neither met on-grade-level proficiency standard nor prepared to be successful at the next level.

Chapter 9 Test Results and Reports

This chapter is divided into two main sections and presents test-level summary results for ***NCEXTEND1*** assessments from Spring 13 operational field tests through the 2014–15 operational administrations. Section one highlights descriptive summary results of raw scores and reported achievement levels for ***NCEXTEND1*** students across major demographic variables. The second section presents samples and summary descriptions of the various standardized reports created by the NCDPI, which are available to LEAs for reporting and interpreting results to stakeholders.

9.1 Raw Score Summary

9.1.1 *NCEXTEND1* Score Distributions

The raw score distributions from the first operational field-test administration of the ***NCEXTEND1*** assessments in 2012–13 are displayed in the score distribution charts in *Figure 9.1* through *Figure 9.7* for ELA, *Figure 9.8* through *Figure 9.14* for math, and *Figure 9.15* through *Figure 9.17* for science. The bar charts indicate that more than half of the students taking the ***NCEXTEND1*** obtained raw scores of 15 or higher in the ***NCEXTEND1*** ELA, Math, and Science.

Figure 9.1 ELA Grade 3 **NCEXTEND1** *Raw Score Distribution 2012–13*

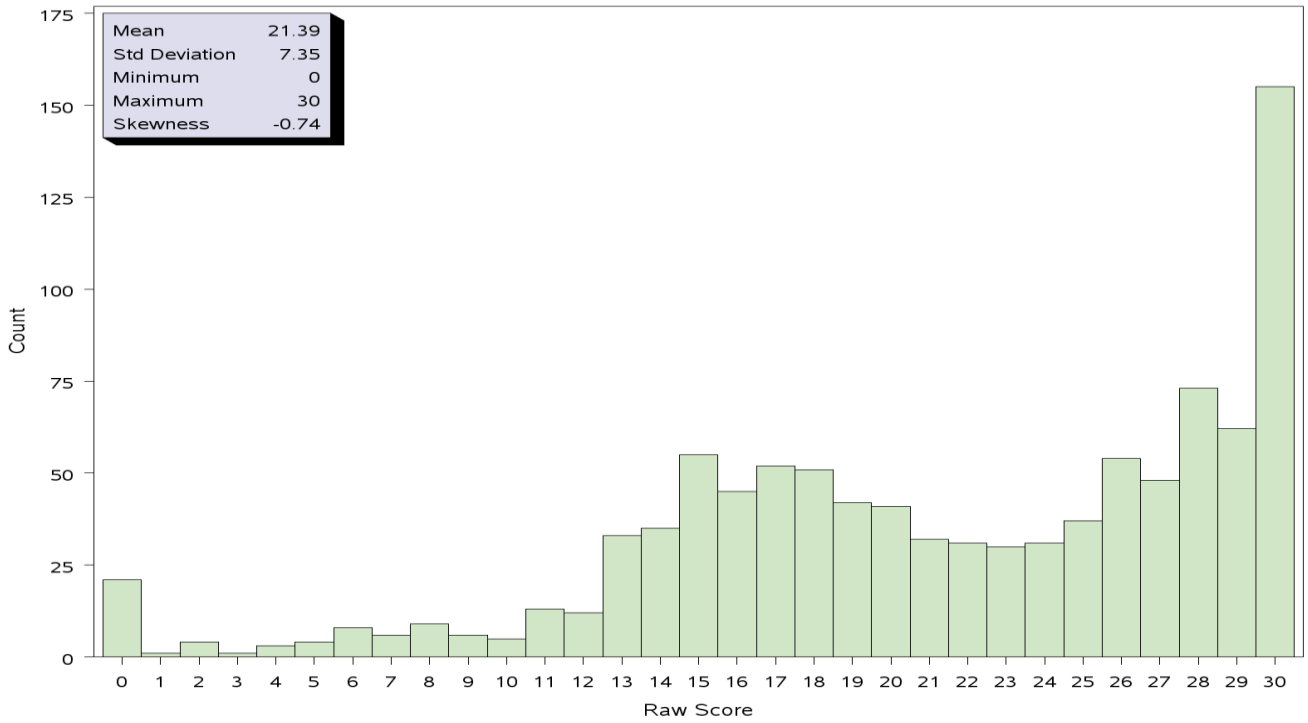


Figure 9.2 ELA Grade 4 **NCEXTEND1** *Raw Score Distribution 2012–13*

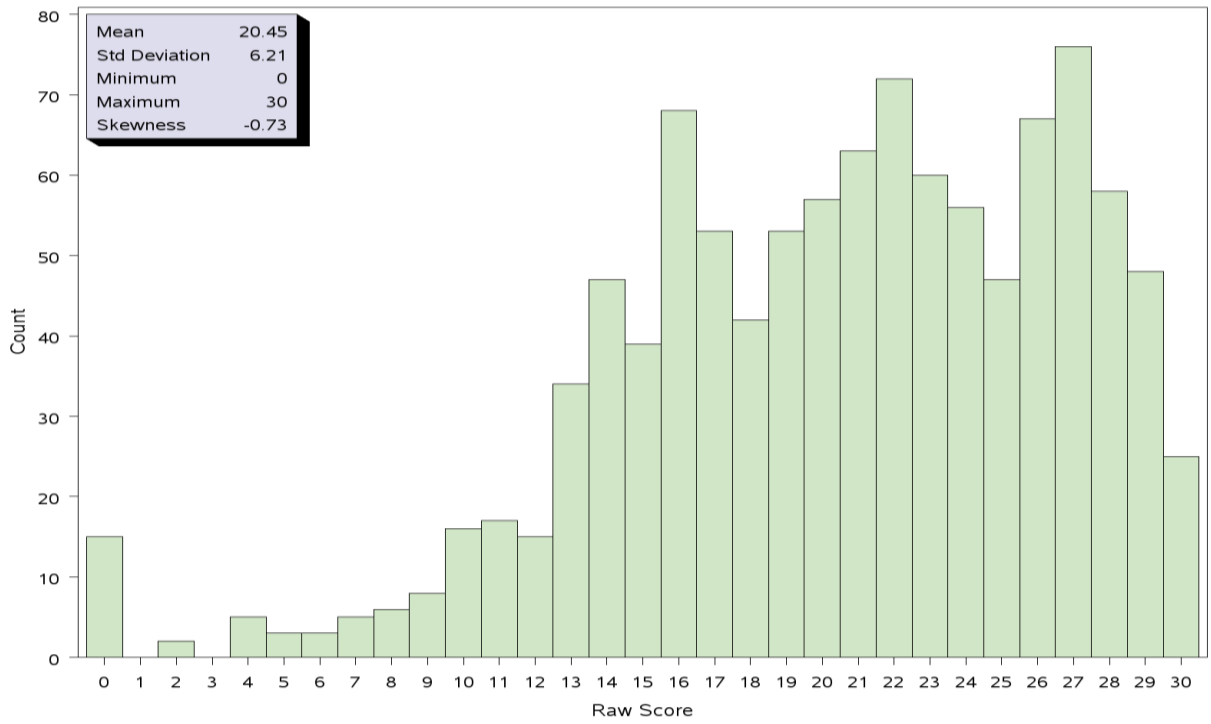


Figure 9.3 ELA Grade 5 **NCEXTEND1** *Raw Score Distribution 2012–13*

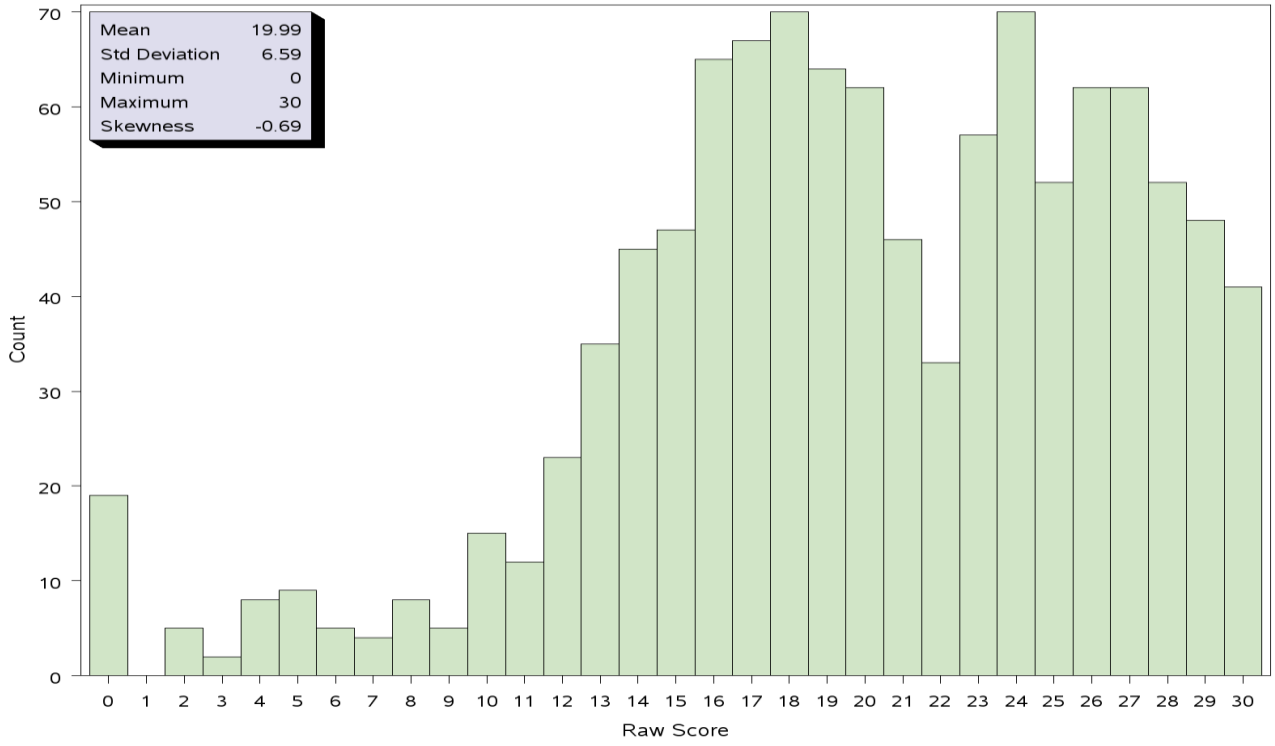


Figure 9.4 ELA Grade 6 **NCEXTEND1** *Raw Score Distribution 2012–13*

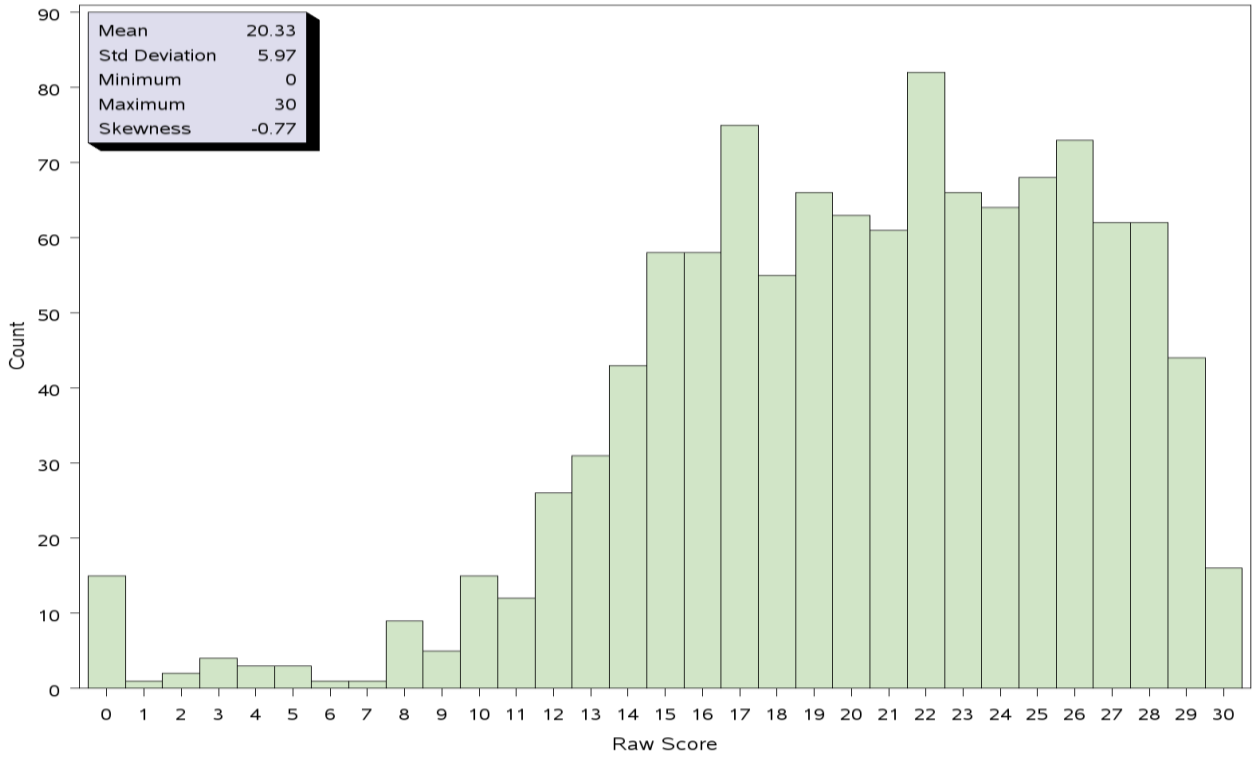


Figure 9.5 ELA Grade 7 **NCEXTEND1** *Scale Score Distribution 2012–13*

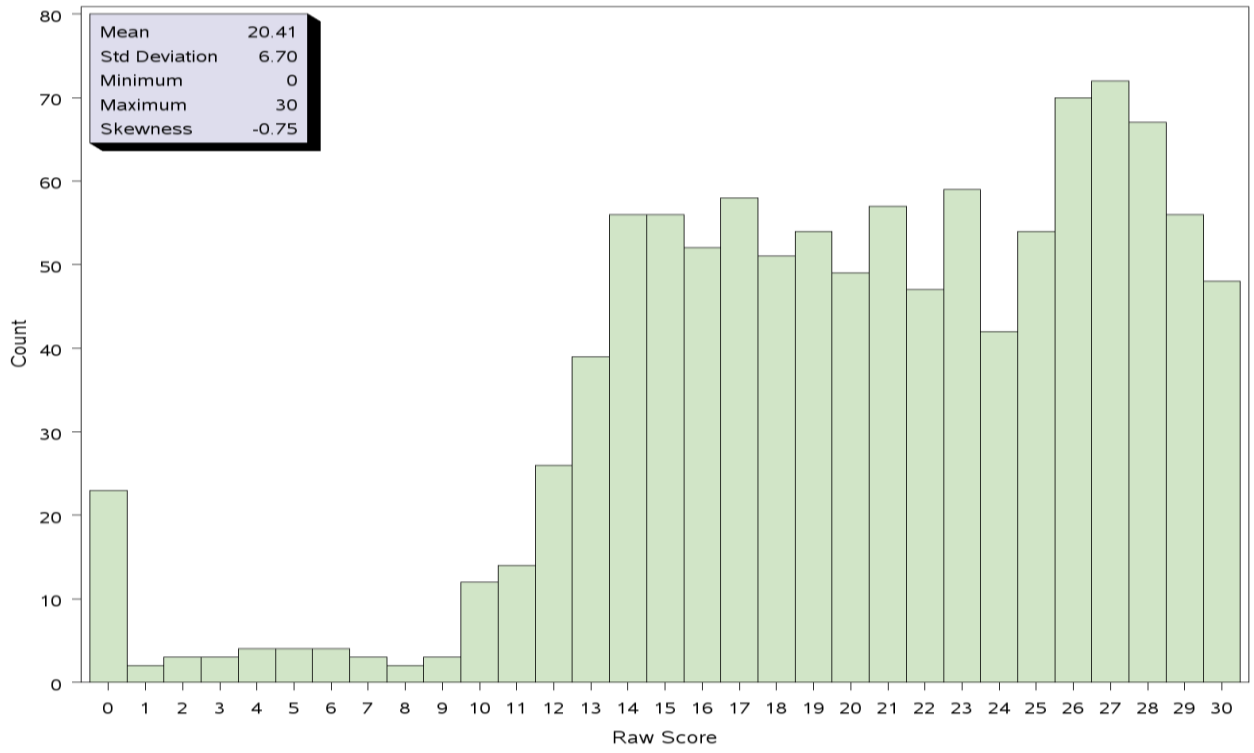


Figure 9.6 ELA Grade 8 **NCEXTEND1** *Raw Score Distribution 2012–13*

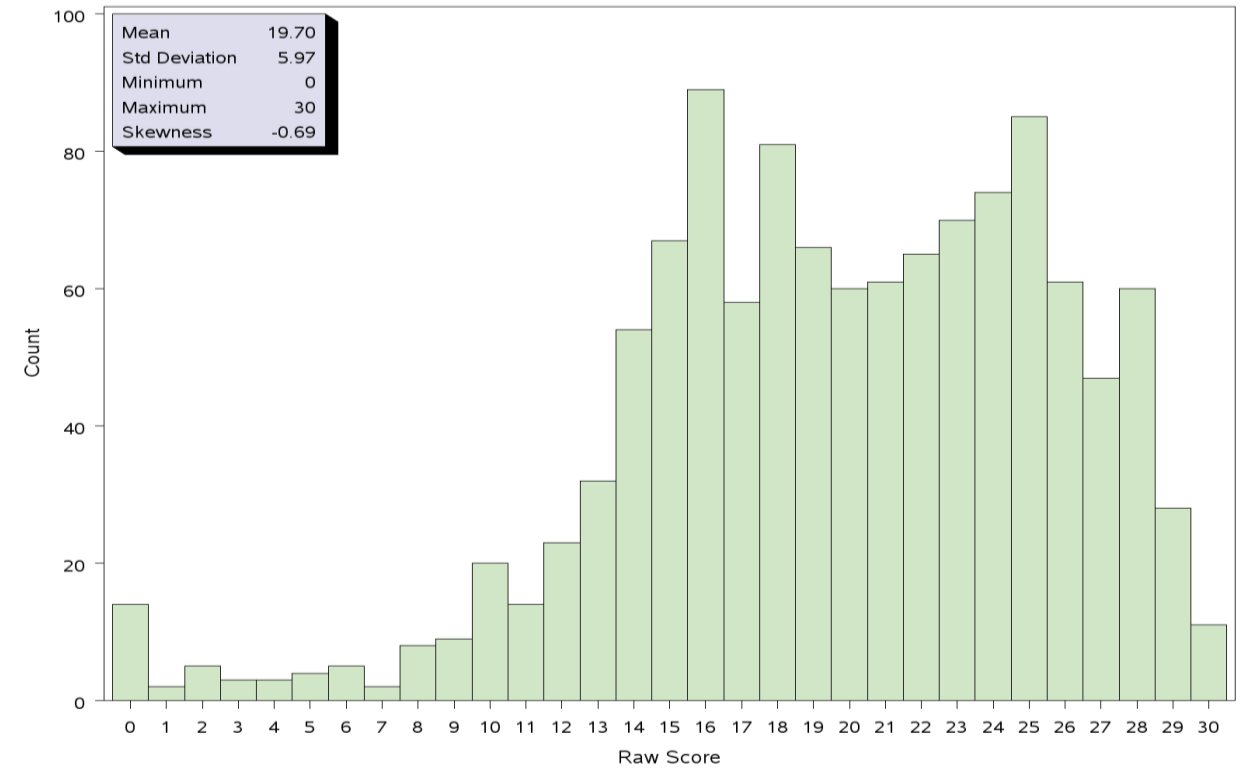


Figure 9.7 English II Grade 10 **NCEXTEND1** *Raw Score Distribution 2012–13*

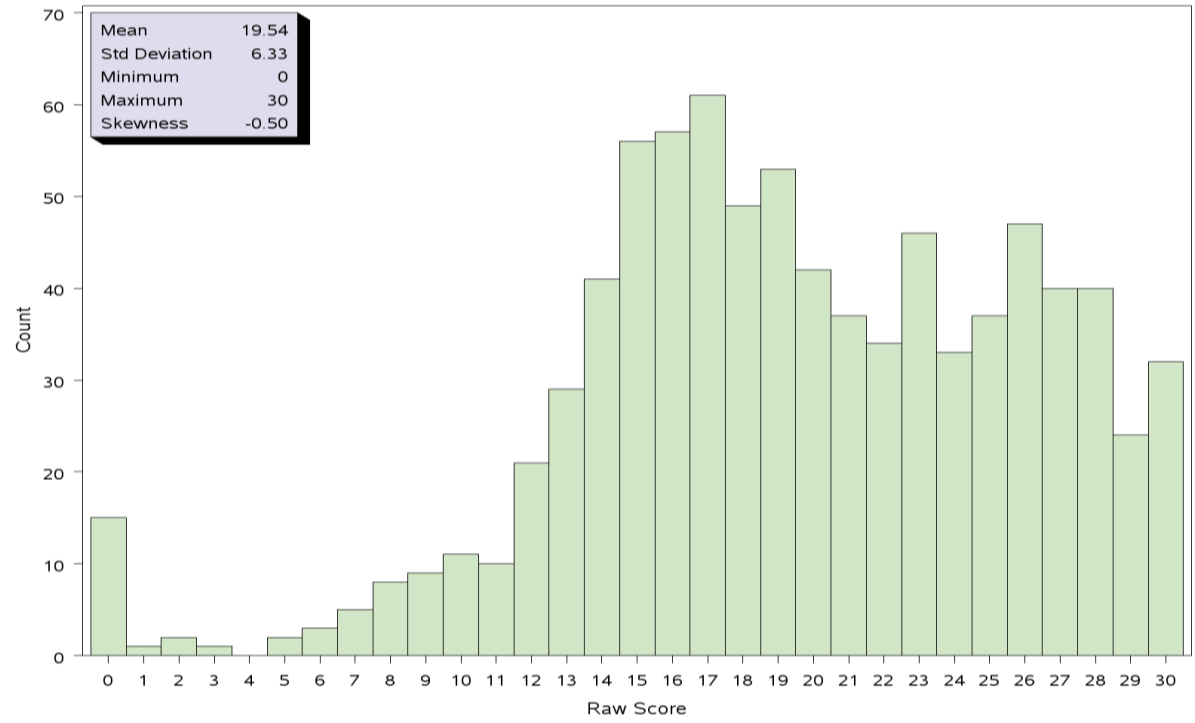


Figure 9.8 Math Grade 3 **NCEXTEND1** *Raw Score Distribution 2012–13*

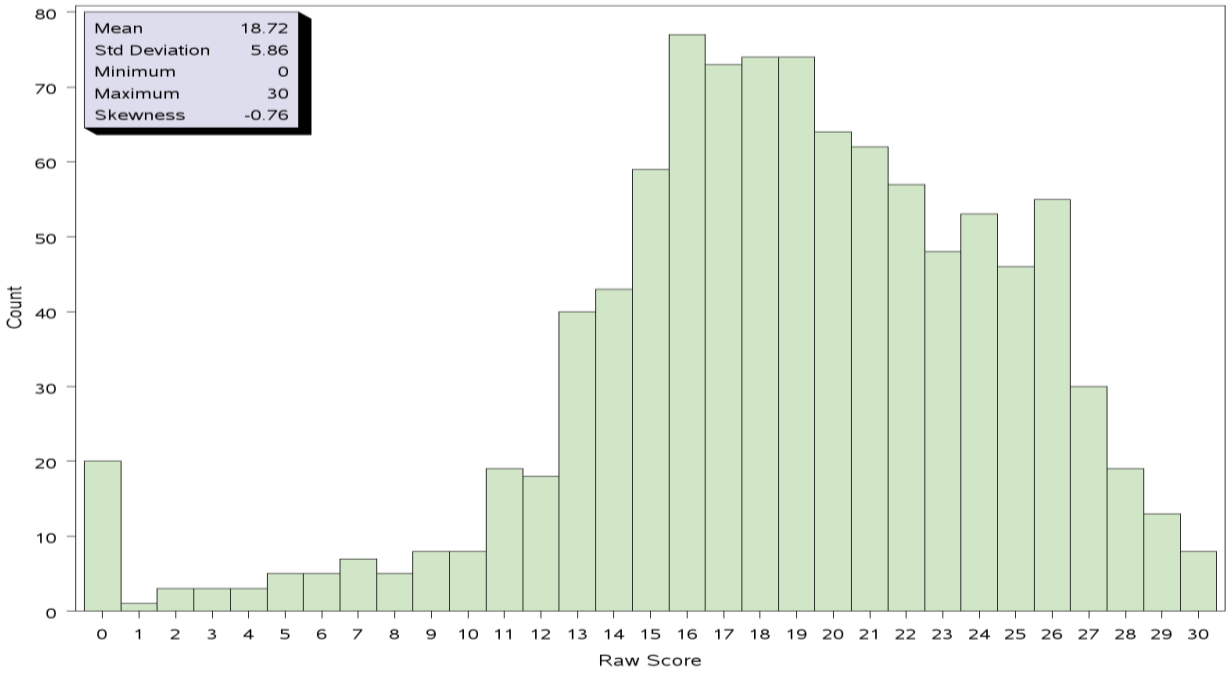


Figure 9.9 Math Grade 4 **NCEXTEND1** *Raw Score Distribution 2012–13*

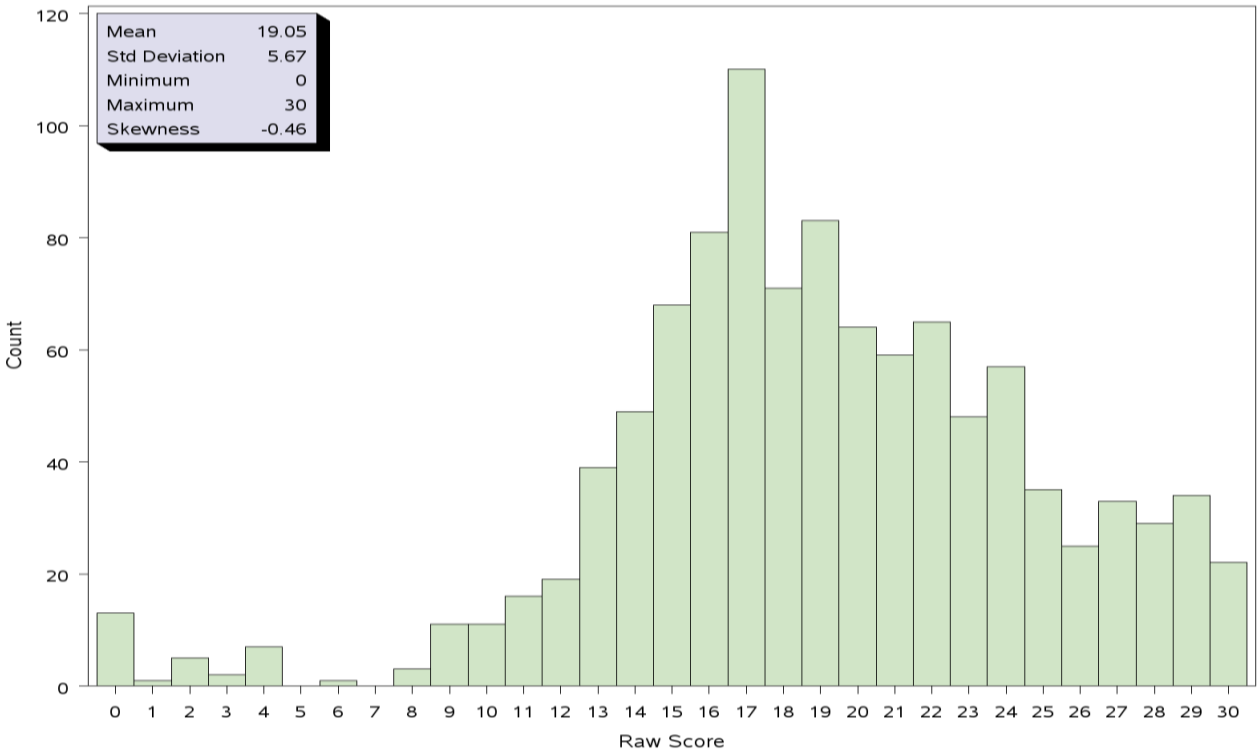


Figure 9.10 Math Grade 5 **NCEXTEND1** *Raw Score Distribution 2012–13*

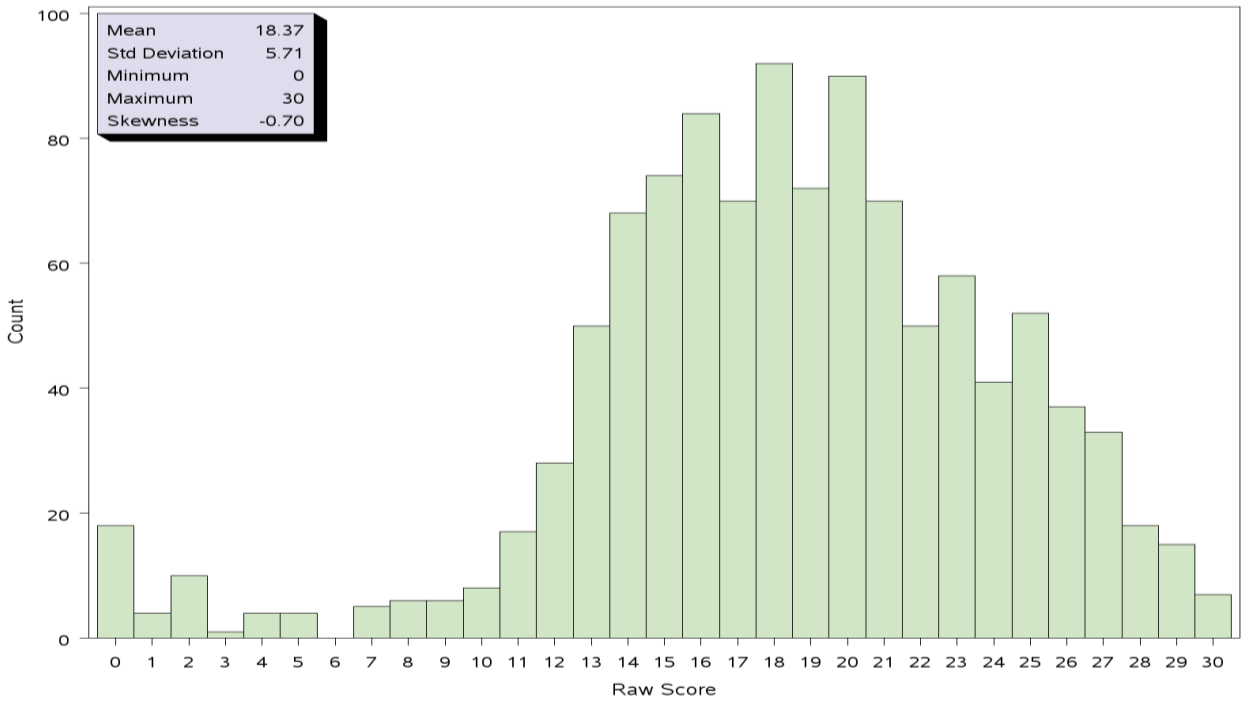


Figure 9.11 Math Grade 6 **NCEXTEND1** *Raw Score Distribution 2012–13*

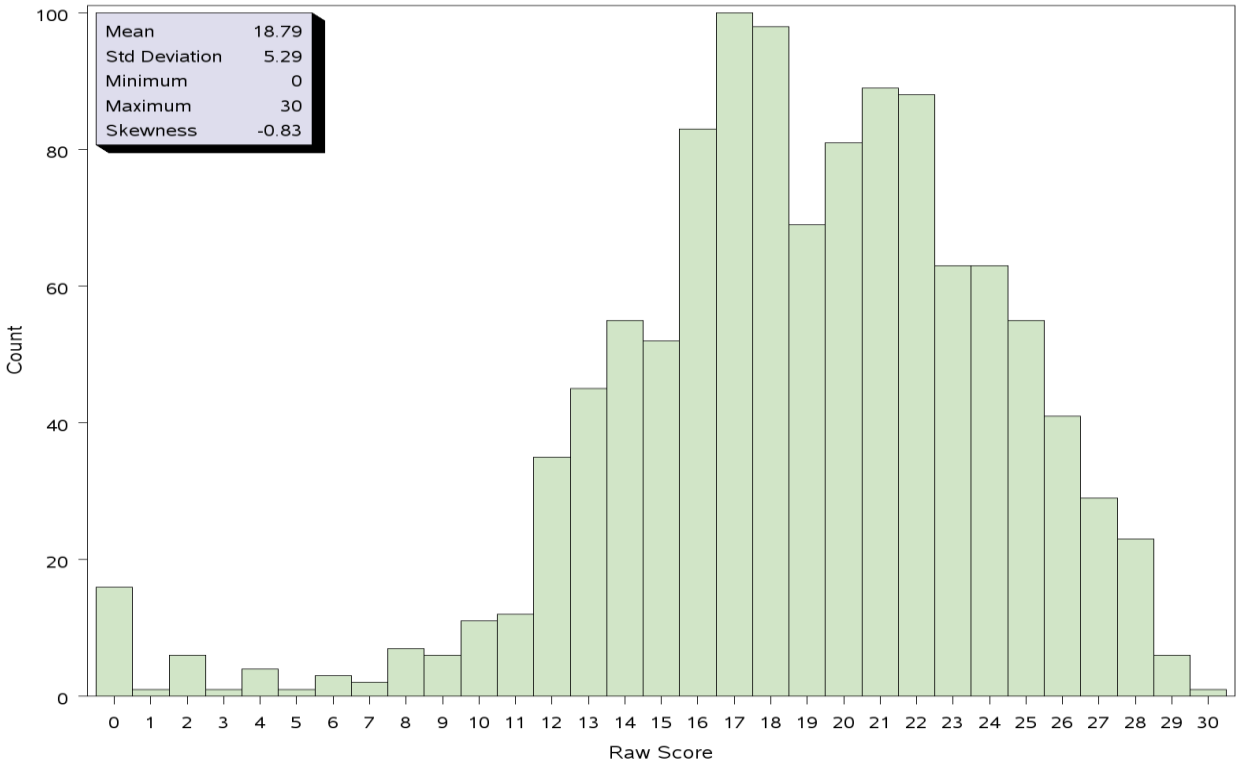


Figure 9.12 Math Grade 7 **NCEXTEND1** *Raw Score Distribution 2012–13*

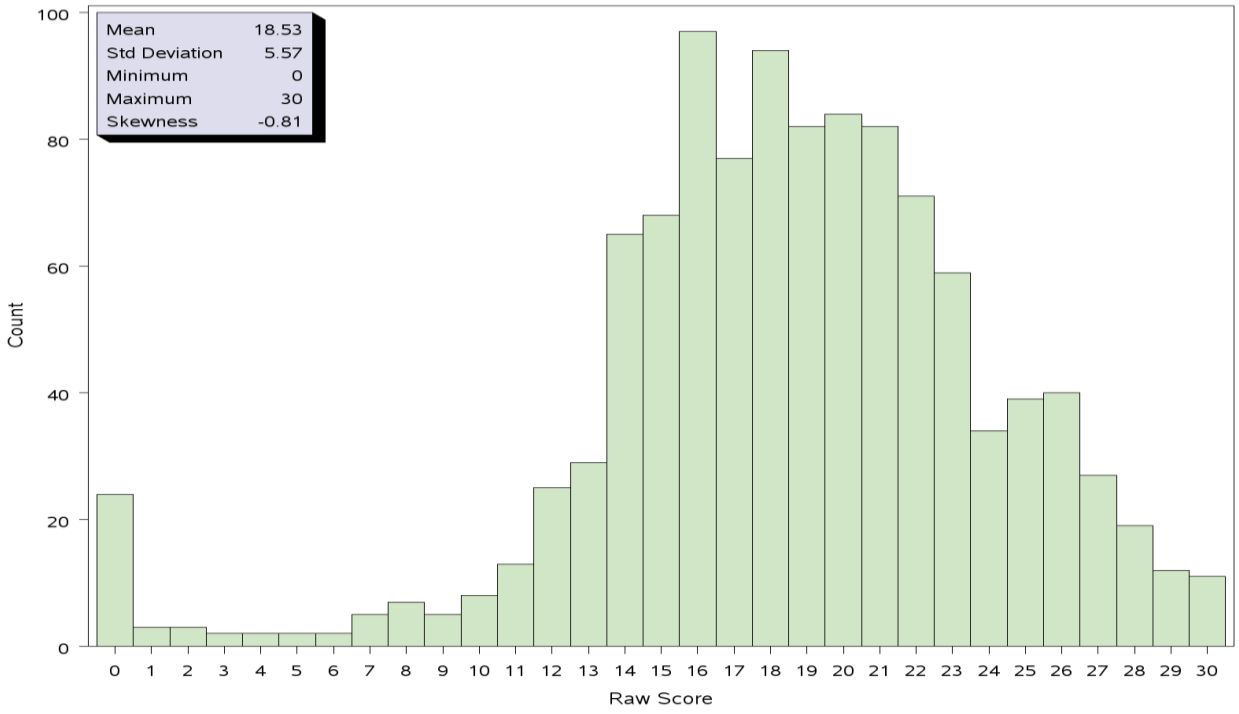


Figure 9.13 Math Grade 8 **NCEXTEND1** *Raw Score Distribution 2012–13*

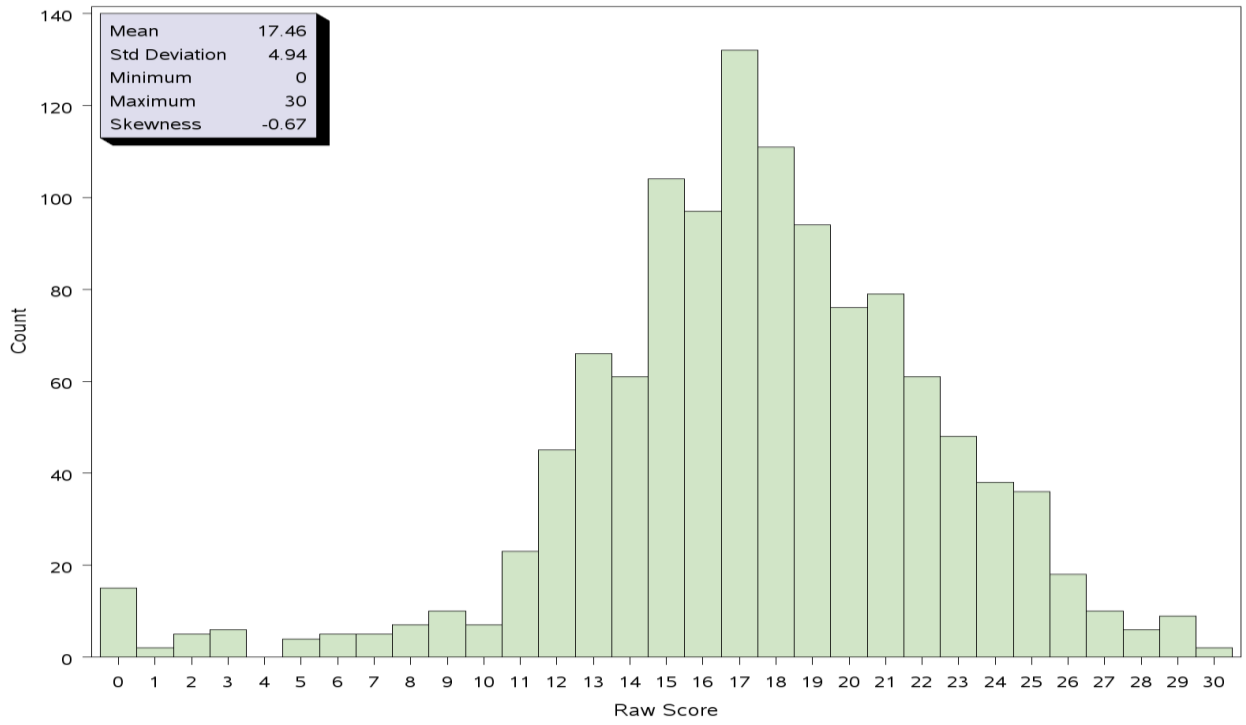


Figure 9.14 Math I Grade 10 **NCEXTEND1** *Raw Score Distribution 2012–13*

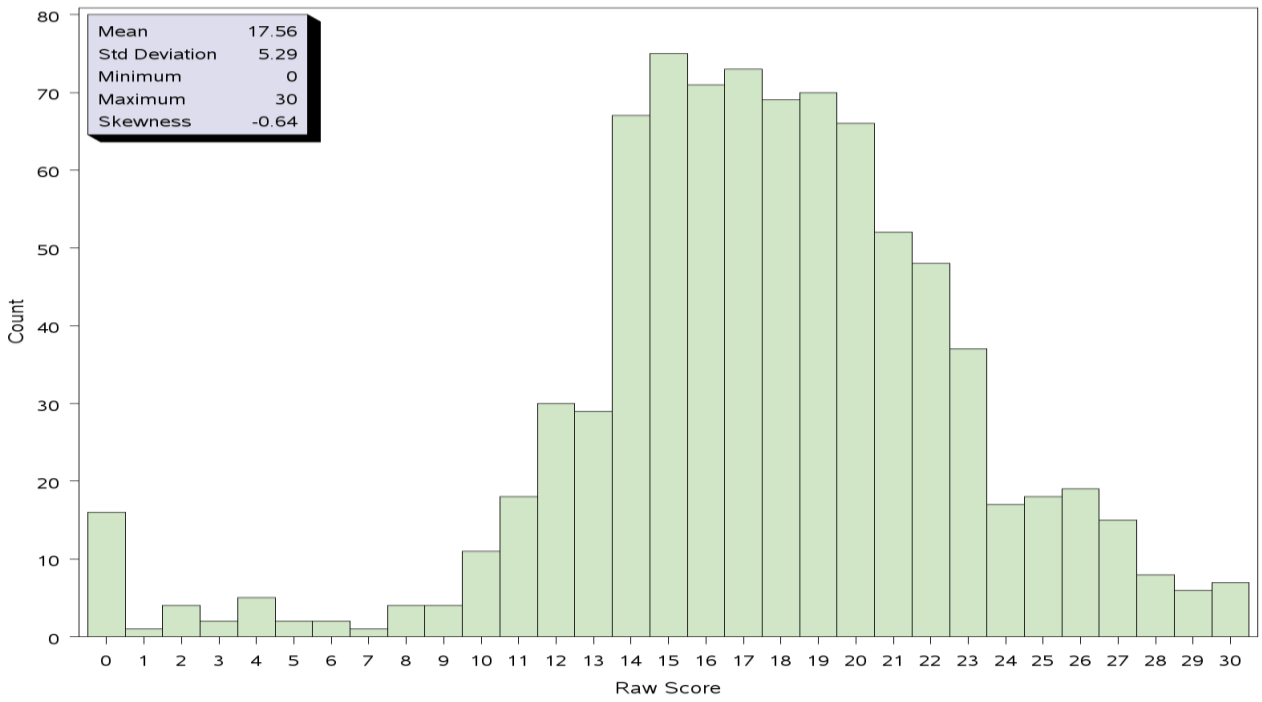


Figure 9.15 Science Grade 5 **NCEXTEND1** *Raw Score Distribution 2012–13*

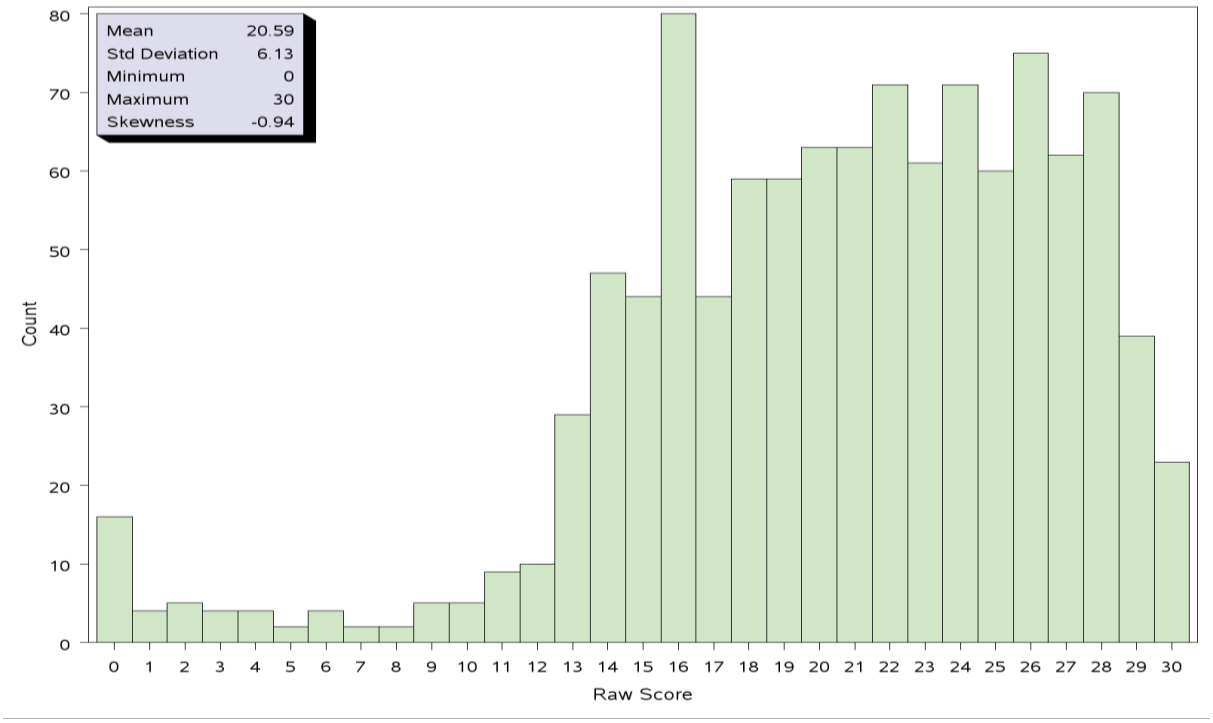


Figure 9.16 Science Grade 8 **NCEXTEND1** *Raw Score Distribution 2012–13*

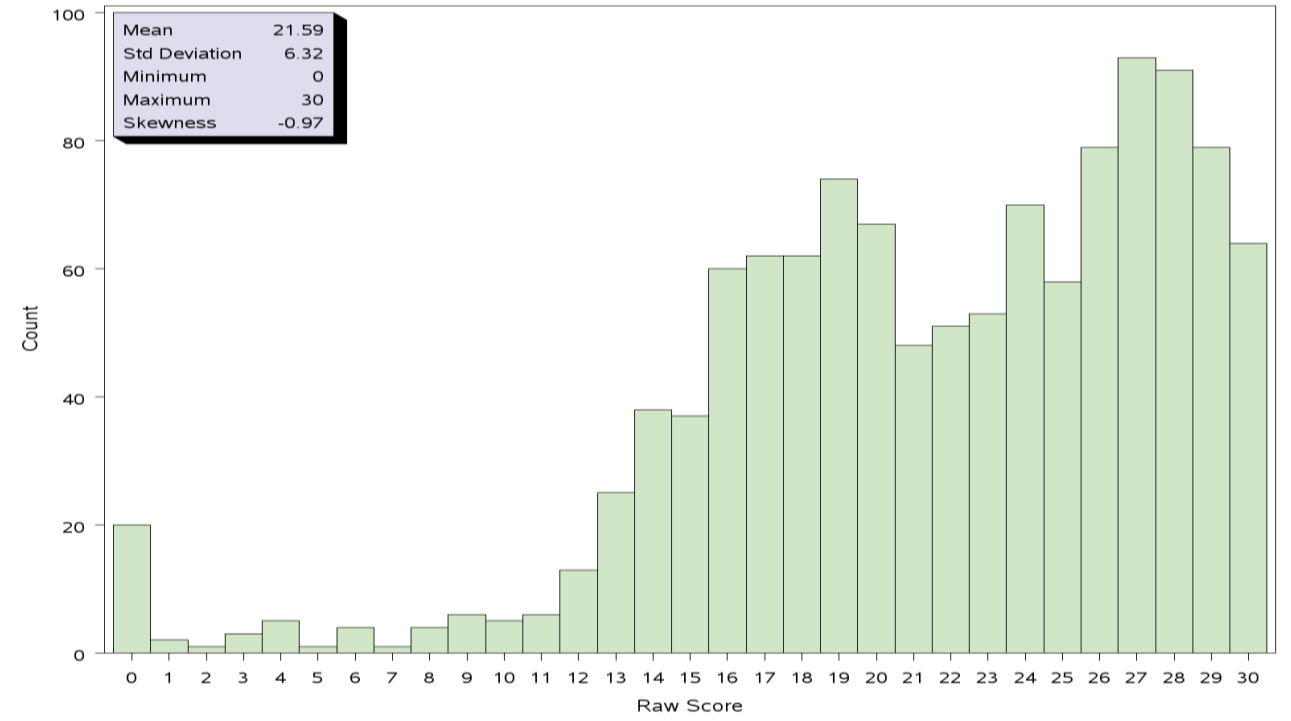
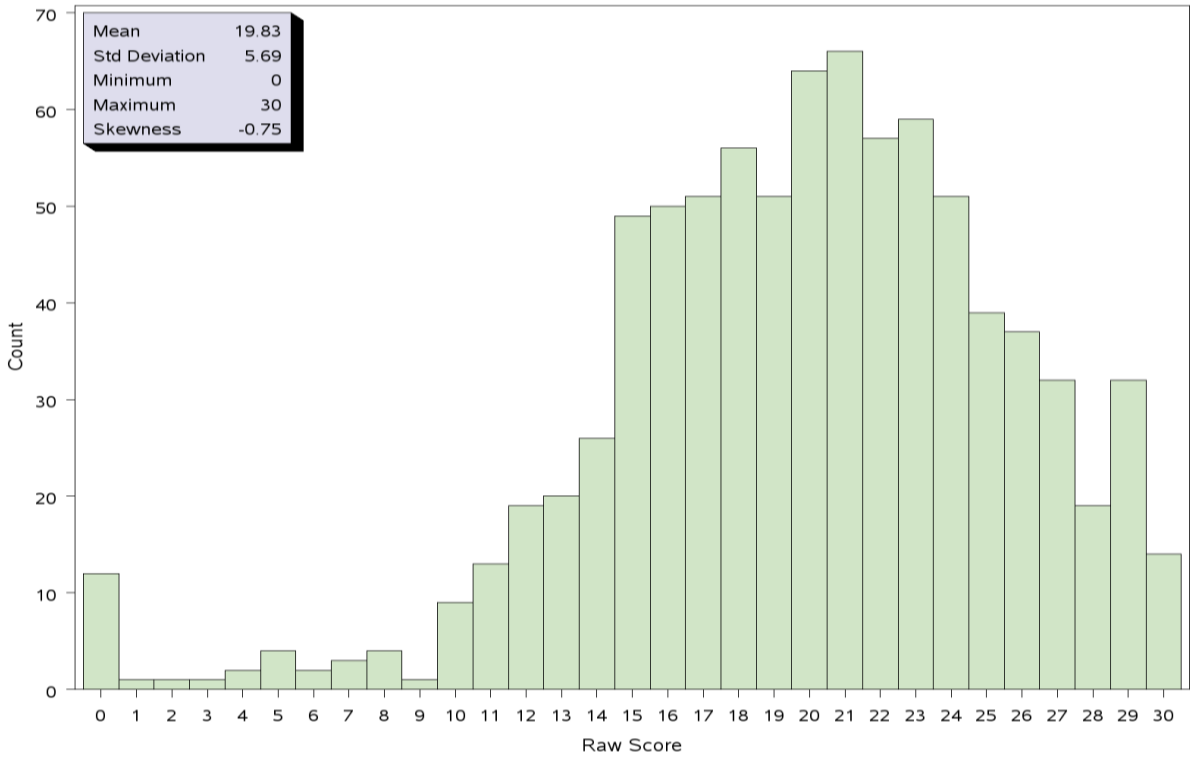


Figure 9.17 Biology Grade 10 **NCEXTEND1** *Raw Score Distribution 2012–13*



A longitudinal summary of ***NCEXTEND1*** raw scores for the most recent three administrations (2012–13, 2013–14, and 2014–15) is presented in *Table 9.1*. The number of students administered ***NCEXTEND1*** assessments across the state has been on a small but steady increase across the years, but overall, ***NCEXTEND1*** populations are still around 1% by grade level (see *Table 5.2* through *Table 5.4*). Descriptive summary evidence from *Table 9.1* indicates mean raw scores have been consistent across the past three years. In general, mean raw scores across all assessments for the past three years have either stayed flat or trended slightly upwards. But the effect of the difference across years is very small and can be mostly explained by sampling variability across years. Overall variability of scores summarized using the standard deviation (SD) also indicates score distribution from 2012–13 to 2014–15 are very similar.

Table 9.1 **NCEXTEND1** *Descriptive Statistics of Scale Scores by Grade across Administrations, Population*

Type	2012–13			2013–14			2014–15		
	N	Mean	SD	N	Mean	SD	N	Mean	SD
ELA									
Grade 3	989	21.4	7.4	1,063	22.1	6.9	1,220	21.9	6.8
Grade 4	1,050	20.5	6.2	1,018	20.5	6.2	1,116	20.7	6.0
Grade 5	1,079	20.0	6.6	1,121	20.4	6.1	1,131	20.2	6.3
Grade 6	1,135	20.3	6.0	1,103	20.4	6.1	1,204	20.6	6.0
Grade 7	1,070	20.4	6.7	1,170	20.7	6.3	1,149	20.9	6.3
Grade 8	1,169	19.8	5.9	1,235	20.1	5.7	1,330	20.4	5.8
Grade 10 English II	834	19.5	6.3	904	20.1	6.3	929	19.5	6.7
Math									
Grade 3	989	18.7	5.9	1,063	19.3	5.7	1,220	19.2	5.3
Grade 4	1,050	19.1	5.6	1,018	19.1	6.0	1,118	19.1	5.6
Grade 5	1,078	18.4	5.7	1,120	18.8	5.2	1,127	18.7	5.4
Grade 6	1,134	18.8	5.3	1,102	19.0	5.4	1,201	19.1	5.2
Grade 7	1,069	18.6	5.6	1,168	18.8	5.3	1,147	19.1	5.3
Grade 8	1,170	17.5	4.9	1,235	17.5	4.7	1,330	18.1	4.8
Grade 10 Math I	835	17.6	5.2	911	18.0	5.0	929	17.7	5.6
Science									
Grade 5	1,078	20.6	6.1	1,114	21.1	5.7	1,128	20.9	6.0
Grade 8	1,169	21.7	6.2	1,234	22.0	6.0	1,329	22.3	5.8
Grade 10 Biology	835	19.9	5.6	911	20.2	5.7	928	19.9	6.1

9.1.2 Raw Score by Gender

Raw score summaries by gender for the ***NCEXTEND1*** scores across three administrations also show similar trends observed in the population distribution. Across all grades, the number of male students is almost double that of female students, with male students scoring slightly higher than female students in most cases. Raw score variance was very similar in both gender groups and has been consistent (see *Table 9.2* through *Table 9.4)* across years.

Table 9.2 ELA **NCEXTEND1** *Raw Scores by Grade and Gender, Population*

ELA	Gender	2012–13			2013–14			2014–15		
		N	Mean	SD	N	Mean	SD	N	Mean	SD
Grade 3	Female	332	21.3	7.3	331	22.0	6.8	401	21.9	6.6
	Male	657	21.4	7.4	732	22.2	7.0	819	21.9	6.9
Grade 4	Female	344	20.5	6.2	339	20.6	6.2	355	20.5	5.9
	Male	706	20.5	6.2	679	20.5	6.1	761	20.7	6.1
Grade 5	Female	332	19.5	6.7	368	20.3	6.1	376	19.9	6.8
	Male	747	20.2	6.5	753	20.4	6.1	755	20.4	6.1
Grade 6	Female	395	19.9	6.2	332	20.2	6.1	414	20.4	5.9
	Male	740	20.6	5.9	771	20.5	6.1	790	20.7	6.1
Grade 7	Female	388	20.5	7.0	390	20.4	6.6	351	20.7	6.4
	Male	682	20.4	6.6	780	20.8	6.1	798	21.0	6.3
Grade 8	Female	396	19.6	5.8	453	20.1	5.6	452	20.2	5.9
	Male	773	19.8	5.9	782	20.1	5.8	878	20.5	5.7
Grade 10 English II	Female	288	19.4	6.2	326	19.7	6.3	310	18.9	6.5
	Male	546	19.6	6.4	578	20.2	6.3	619	19.8	6.8

Table 9.3 Math **NCEXTEND1** *Raw Scores by Grade and Gender, Population*

Math	Gender	2012–13			2013–14			2014–15		
		N	Mean	SD	N	Mean	SD	N	Mean	SD
Grade 3	Female	332	18.3	5.7	331	18.8	5.5	401	19.1	5.0
	Male	657	18.9	5.9	732	19.5	5.8	819	19.2	5.4
Grade 4	Female	343	18.6	5.6	339	18.7	6.0	358	18.6	5.6
	Male	707	19.3	5.6	679	19.2	5.9	760	19.4	5.6
Grade 5	Female	332	17.8	5.7	367	18.1	5.3	375	18.2	5.7
	Male	746	18.7	5.7	753	19.1	5.1	752	18.9	5.2
Grade 6	Female	394	18.6	5.4	330	18.8	5.3	414	18.7	5.0
	Male	740	18.9	5.3	772	19.1	5.4	787	19.4	5.3
Grade 7	Female	386	18.2	5.5	390	18.2	5.3	350	18.7	5.1
	Male	683	18.8	5.6	778	19.0	5.3	797	19.3	5.4
Grade 8	Female	396	17.2	4.8	453	17.2	4.4	452	17.8	4.9
	Male	774	17.6	5.0	782	17.7	4.9	878	18.2	4.8
Grade 10 Math I	Female	288	17.7	4.9	329	17.6	5.1	310	16.7	5.5
	Male	547	17.6	5.4	582	18.2	5.0	619	18.1	5.5

Table 9.4 Science **NCEXTEND1** *Raw Scores by Grade and Gender, Population*

Science	Gender	2012–13			2013–14			2014–15		
		N	Mean	SD	N	Mean	SD	N	Mean	SD
Grade 5	Female	333	20.2	6.2	365	20.7	5.8	375	20.5	6.3
	Male	745	20.8	6.1	749	21.3	5.6	753	21.1	5.8
Grade 8	Female	396	21.2	6.1	453	22.1	5.9	452	22.3	5.7
	Male	773	21.9	6.3	781	22.0	6.0	877	22.2	5.9
Grade 10 Biology	Female	288	19.7	5.4	329	19.8	5.8	310	19.2	5.9
	Male	547	20.0	5.7	582	20.4	5.6	618	20.3	6.2

9.1.3 Raw Score by Disability

The mean raw scores by various sub-groups are shown in *Appendix 9-A* ***NCEXTEND1*** *Raw Score by Subgroups*. The data indicated that the number of accommodated students are not similar across administrations for some categories. The mean raw scores are similar across administrations for the most accommodated groups. Note that only accommodation categories with five or more students are reported. There are many categories where the number of students are less than five in some administrations.

9.2 Achievement Levels

The achievement level classifications for the population across grades and administrations are displayed in *Table 9.5* through *Table 9.7*. Note that the cut scores for the base administration (2012–13) were different from the 2013–14 administration and beyond; and as a result in 2012–13, NCDPI classified students using 4 achievement levels. From 2013–14 onwards students are classified based on a 5-achievement-level scale. Therefore, achievement level proportions for 2012–13 cannot be directly compared with those from subsequent administrations. For 2013–14 administration and beyond Level 3 "Sufficient Command" was added, and Levels 3 and 4 became Levels 4 and 5 respectively. For 2012–13 administration in the *Table 9.5* through *Table 9.7* there is no data for Level 3. Levels 3 and 4 proportions for 2012–13 have been displayed as Levels 4 and 5 respectively. The short-term trend between 2013–14 and 2014–15 varies across these three subjects.

Achievement levels distributions for ELA shows about a 2% increase in the proportion of students classified as meeting on grade proficiency and prepared to be successful at the next level (Levels 4 and 5) for grades 4, 6, 7, and 8 between 2013–14 and 2014–15. For grades 3, 5 and 10, the proportion of students in Levels 4 and 5 has actually decreased by 3%.

In math, the proportion of students in Levels 4 and 5 is either staying constant or trending upwards in grades 5, 7, 8 and grade 10 by as little as 0.3% to up to 4%. The proportion decreased by over 4.7% in grades 3, and about 1% in grades 4 and 6.

In Science, 40% or more of students were classified as Levels 4 and 5 across all three grades.

Table 9.5 ELA **NCEXTEND1** *Achievement level classifications by Grade and Year*

ELA	Year	N	*% Achievement Level*				
			1) Limited Command	*2) Partial Command*	*3) Sufficient Command*	*4) Solid Command*	*5) Superior Command*
Grade 3	2012–13*	989	31.5	19.5	*	27.3	21.7
	2013–14	1,063	18.3	21.3	7.2	28.4	24.7
	2014–15	1,220	18.3	23.6	6.6	29.4	22.1
Grade 4	2012–13*	1,050	26.7	32	*	21.8	19.5
	2013–14	1,018	10.5	31.5	16.6	21.2	20.1
	2014–15	1,116	9.7	32.4	14.5	23.5	20
Grade 5	2012–13*	1,079	27.9	18.4	*	29.4	24.4
	2013–14	1,121	11.4	15.5	16.2	32	24.8
	2014–15	1,131	12.1	17.2	17	26.8	27
Grade 6	2012–13*	1,135	31.5	28.7	*	29	10.8
	2013–14	1,103	15.1	28	16.5	28.4	12
	2014–15	1,204	15	26.1	16.1	30.3	12.5
Grade 7	2012–13*	1,070	33.1	28.9	*	9	29.1
	2013–14	1,170	15.7	33.4	11.4	10.3	29.2
	2014–15	1,149	16.1	29.3	13.6	9.8	31.2
Grade 8	2012–13*	1,169	34.7	28.2	*	24.6	12.5
	2013–14	1,235	15.2	27.5	18.1	27.3	11.9
	2014–15	1,330	14.4	28.4	16.1	26.5	14.7
Grade 10	2012–13*	834	32.1	24.3	*	22.1	21.5
	2013–14	904	14.4	20	18.8	22.5	24.3
	2014–15	929	16.3	22.8	16.8	22.7	21.4

**The SBE adopted four achievement levels for 2012–13, hence the results are not comparable with 2013–14 and 2014–15.*

Table 9.6 Math NCEXTEND1 *Achievement level classifications by Grade and Year*

Math	Year	N	% Achievement Level				
			1) Limited Command	*2) Partial Command*	*3) Sufficient Command*	*4) Solid Command*	*5) Superior Command*
Grade 3	2012–13*	989	32.6	40	*	23.36	4
	2013–14	1,063	10.7	39.9	17.7	25.2	6.5
	2014–15	1,220	9.4	42.4	21.2	22.8	4.2
Grade 4	2012–13*	1,050	30.8	30.9	*	27.3	11.1
	2013–14	1,018	14.9	27	16.6	29.7	11.8
	2014–15	1,118	12.5	27.7	19.2	30.5	10
Grade 5	2012–13*	1,078	35.3	36.1	*	18.5	10.2
	2013–14	1,120	11.7	35.6	23.8	19.6	9.3
	2014–15	1,127	11.9	37.6	19.9	20.1	10.5
Grade 6	2012–13*	1,134	29.5	45.9	*	21.9	2.7
	2013–14	1,102	12.3	37.8	22	25.1	2.9
	2014–15	1,201	11.6	39.3	22.2	24.2	2.8
Grade 7	2012–13*	1,069	39.7	46.6	*	11.6	2.2
	2013–14	1,168	17	54.1	15.1	10.8	3.1
	2014–15	1,147	16.4	50.5	18	12.6	2.6
Grade 8	2012–13*	1,170	50.4	35.4	*	12.7	1.5
	2013–14	1,235	21.2	46.4	19.4	11.1	1.9
	2014–15	1,330	18.5	43.8	20.1	15.6	2
Grade 10	2012–13*	835	40.4	25.2	*	25.9	8.6
	2013–14	911	14.2	23.4	25.3	27.2	10
	2014–15	929	16.2	22.6	23.8	28.1	9.4

**The SBE adopted four achievement levels for 2012–13, hence the results are not comparable with 2013–14 and 2014–15.*

Table 9.7 Science **NCEXTEND1** *Achievement level classifications by Grade and Year*

Science	Year	N	% Achievement Level				
			1) Limited Command	*2) Partial Command*	*3) Sufficient Command*	*4) Solid Command*	*5) Superior Command*
Grade 5	2012–13*	1,078	25.1	32.7	*	17.7	24.6
	2013–14	1,114	7.6	29.4	17.3	19.8	25.9
	2014–15	1,128	8.4	30.4	16.8	16.8	27.6
Grade 8	2012–13*	1,169	24.5	30.1	*	25.5	19.9
	2013–14	1,234	9.7	28.3	14	27.6	20.4
	2014–15	1,329	9.3	27.2	13.1	29.7	20.7
Grade 10	2012–13*	835	25.6	34.1	*	24.4	15.8
	2013–14	911	8.3	28.2	20	25.9	17.6
	2014–15	928	9.5	28.1	19.8	23.3	19.3

**The SBE adopted four achievement levels for 2012–13, hence the results are not comparable with 2013–14 and 2014–15.*

Achievement-levels classifications by gender across grades and administrations are shown in *Table 9.8* through *Table 9.12*. These tables should be interpreted with similar caution as the previous table with regards to achievement levels for 2012–13. A similar trend as the total population can be observed between genders. The results across all subjects, administrations and grades show slightly more males than female students scoring at level 4 or above.

Table 9.8 ELA **NCEXTEND1** *Achievement level classifications by Gender Grades 3–8*

ELA	Year	Gender	N	Achievement Level				
				1) Limited Command	*2) Partial Command*	*3) Sufficient Command*	*4) Solid Command*	*5) Superior Command*
Grade 3	2012–13*	Female	332	29.2	22.9	*	26.8	21.1
		Male	657	32.6	17.8	*	27.6	22.1
	2013–14	Female	331	17.8	22.4	7.9	28.4	23.6
		Male	732	18.6	20.8	7	28.4	25.3
	2014–15	Female	401	17.5	24.9	6	30.2	21.5
		Male	819	18.7	23	6.8	29.1	22.5
Grade 4	2012–13*	Female	344	25.9	33.7	*	21.5	18.9
		Male	706	27.1	31.2	*	22	19.8
	2013–14	Female	339	10.6	30.7	19.5	17.7	21.5
		Male	679	10.5	32	15.2	23	19.4
	2014–15	Female	355	10.4	32.1	14.4	25.9	17.2
		Male	761	9.3	32.5	14.6	22.3	21.3
Grade 5	2012–13*	Female	332	28.3	19.9	*	31	20.8
		Male	747	27.7	17.7	*	28.7	26
	2013–14	Female	368	12	12.2	17.7	35.3	22.8
		Male	753	11.2	17.1	15.5	30.4	25.8
	2014–15	Female	376	14.1	14.9	15.4	30.1	25.5
		Male	755	11.1	18.3	17.8	25.2	27.7
Grade 6	2012–13*	Female	395	33.4	29.6	*	27.1	9.9
		Male	740	30.5	28.2	*	30	11.2
	2013–14	Female	332	15.4	28	18.4	25.9	12.4
		Male	771	15.1	28	15.7	29.4	11.8
	2014–15	Female	414	15.7	25.9	16.7	31.4	10.4
		Male	790	14.7	26.2	15.8	29.8	13.5
Grade 7	2012–13*	Female	388	34.8	24.2	*	10.1	30.9
		Male	682	32.1	31.5	*	8.4	28
	2013–14	Female	390	17.2	32.6	12.3	11	26.9
		Male	780	15	33.9	10.9	10	30.3
	2014–15	Female	351	17.1	30.5	11.4	10.5	30.5
		Male	798	15.7	28.8	14.5	9.5	31.5
Grade 8	2012–13*	Female	396	34.1	30.8	*	22.7	12.4
		Male	773	35.1	26.9	*	25.5	12.6
	2013–14	Female	453	16.3	23.8	22.3	27.4	10.2
		Male	782	14.6	29.5	15.7	27.2	12.9
	2014–15	Female	452	15.5	28.3	15	27.9	13.3

**The SBE adopted four achievement levels for 2012–13, hence the results are not comparable with 2013–14 and 2014–15.*

*Table 9.9 English II **NCEXTEND1** Achievement level classifications by Gender Grade 10*

ELA	Year	Gender	N	Achievement Level				
				1) Limited Command	*2) Partial Command*	*3) Sufficient Command*	*4) Solid Command*	*5) Superior Command*
		Male	819	18.7	23	6.8	29.1	22.5
Grade 10	2012–13*	Female	288	34	24.3	*	20.5	21.2
		Male	546	31.1	24.4	*	22.9	21.6
	2013–14	Female	326	15	18.7	21.8	22.7	21.8
		Male	578	14	20.8	17.1	22.3	25.8
	2014–15	Female	310	17.7	23.2	19	23.2	16.8
		Male	619	15.5	22.6	15.7	22.5	23.8

**The SBE adopted four achievement levels for 2012–13, hence the results are not comparable with 2013–14 and 2014–15.*

*Table 9.10 Math **NCEXTEND1** Achievement level classifications by Gender*

Grade	Year	Gender	N	Achievement Level				
				1) Limited Command	*2) Partial Command*	*3) Sufficient Command*	*4) Solid Command*	*5) Superior Command*
Grade 3	2012–13*	Female	332	34	42.2	*	21.4	2.4
		Male	657	31.8	39	*	24.4	4.9
	2013–14	Female	331	10.9	42.9	18.1	24.5	3.6
		Male	732	10.7	38.5	17.5	25.6	7.8
	2014–15	Female	401	10	43.1	22.4	20.2	4.2
		Male	819	9.2	42	20.6	24.1	4.2
Grade 4	2012–13*	Female	343	32.4	30.9	*	28.6	8.2
		Male	707	30	30.8	*	26.7	12.5
	2013–14	Female	339	16.8	29.2	15.3	27.1	11.5
		Male	679	14	25.9	17.2	30.9	11.9
	2014–15	Female	358	14.3	28.2	21.2	28.8	7.5
		Male	760	11.7	27.5	18.3	31.3	11.2
Grade 5	2012–13*	Female	332	36.1	40.1	*	16.9	6.9
		Male	746	34.9	34.3	*	19.2	11.7
	2013–14	Female	367	14.4	36	24.3	17.7	7.6
		Male	753	10.4	35.5	23.5	20.6	10.1
	2014–15	Female	375	12.3	37.9	22.9	19.7	7.2
		Male	752	11.7	37.5	18.4	20.4	12.1
Grade 6	2012–13*	Female	394	28.2	50.5	*	18.3	3.1
		Male	740	30.3	43.5	*	23.8	2.4
	2013–14	Female	330	10.3	41.5	23	23	2.1
		Male	772	13.2	36.1	21.5	25.9	3.2
	2014–15	Female	414	13	43.2	21.3	20.5	1.9
		Male	787	10.8	37.2	22.6	26.1	3.3
Grade 7	2012–13*	Female	386	40.9	48.5	*	9.3	1.3
		Male	683	39	45.5	*	12.9	2.6
	2013–14	Female	390	18.7	56.2	13.9	8.7	2.6
		Male	778	16.1	53.1	15.7	11.8	3.3
	2014–15	Female	350	18.3	52.3	17.7	10.9	0.9
		Male	797	15.6	49.7	18.1	13.3	3.4
Grade 8	2012–13*	Female	396	53.5	32.3	*	12.6	1.5
		Male	774	48.8	37	*	12.8	1.4
	2013–14	Female	453	19.4	54.5	16.3	8.6	1.1
		Male	782	22.3	41.7	21.1	12.5	2.4
	2014–15	Female	452	19.3	46.2	19.7	13.1	1.8

**The SBE adopted four achievement levels for 2012–13, hence the results are not comparable with 2013–14 and 2014–15.*

*Table 9.11 Math I **NCEXTEND1** Achievement level classifications by Gender*

Grade	Year	Gender	N	Achievement Level				
				1) Limited Command	*2) Partial Command*	*3) Sufficient Command*	*4) Solid Command*	*5) Superior Command*
Grade 10	2012–13*	Female	288	38.5	26.4	*	27.4	7.6
		Male	547	41.3	24.5	*	25.1	9.1
	2013–14	Female	329	14.9	25.5	24.3	26.4	8.8
		Male	582	13.8	22.2	25.8	27.7	10.7
	2014–15	Female	310	20.3	25.2	24.8	23.6	6.1
		Male	619	14.1	21.3	23.3	30.4	11

**The SBE adopted four achievement levels for 2012–13, hence the results are not comparable with 2013–14 and 2014–15.*

*Table 9.12 Science **NCEXTEND1** Achievement level classifications by Gender*

Grade	Year	Gender	N	Achievement Level				
				1) Limited Command	*2) Partial Command*	*3) Sufficient Command*	*4) Solid Command*	*5) Superior Command*
Grade 5	2012–13*	Female	333	23.4	36	*	20.4	20.1
		Male	745	25.8	31.1	*	16.5	26.6
	2013–14	Female	365	7.7	30.4	17	23.6	21.4
		Male	749	7.6	28.8	17.5	18	28
	2014–15	Female	375	9.3	28.3	18.4	20.8	23.2
		Male	753	8	31.5	16.1	14.7	29.8
Grade 8	2012–13*	Female	396	25.8	34.1	*	23.5	16.7
		Male	773	23.8	28.1	*	26.5	21.6
	2013–14	Female	453	8	30.2	13.9	28.7	19.2
		Male	781	10.8	27.1	14.1	26.9	21.1
	2014–15	Female	452	7.7	26.8	15.3	32.1	18.1
		Male	877	10.2	27.5	12	28.4	22
Grade 10	2012–13*	Female	288	24.7	39.2	*	22.6	13.5
		Male	547	26.1	31.4	*	25.4	17
	2013–14	Female	329	9.7	28	21.3	26.1	14.9
		Male	582	7.6	28.4	19.2	25.8	19.1
	2014–15	Female	310	10.7	29.4	24.8	22.6	12.6
		Male	618	8.9	27.5	17.3	23.6	22.7

**The SBE adopted four achievement levels for 2012–13, hence the results are not comparable with 2013–14 and 2014–15.*

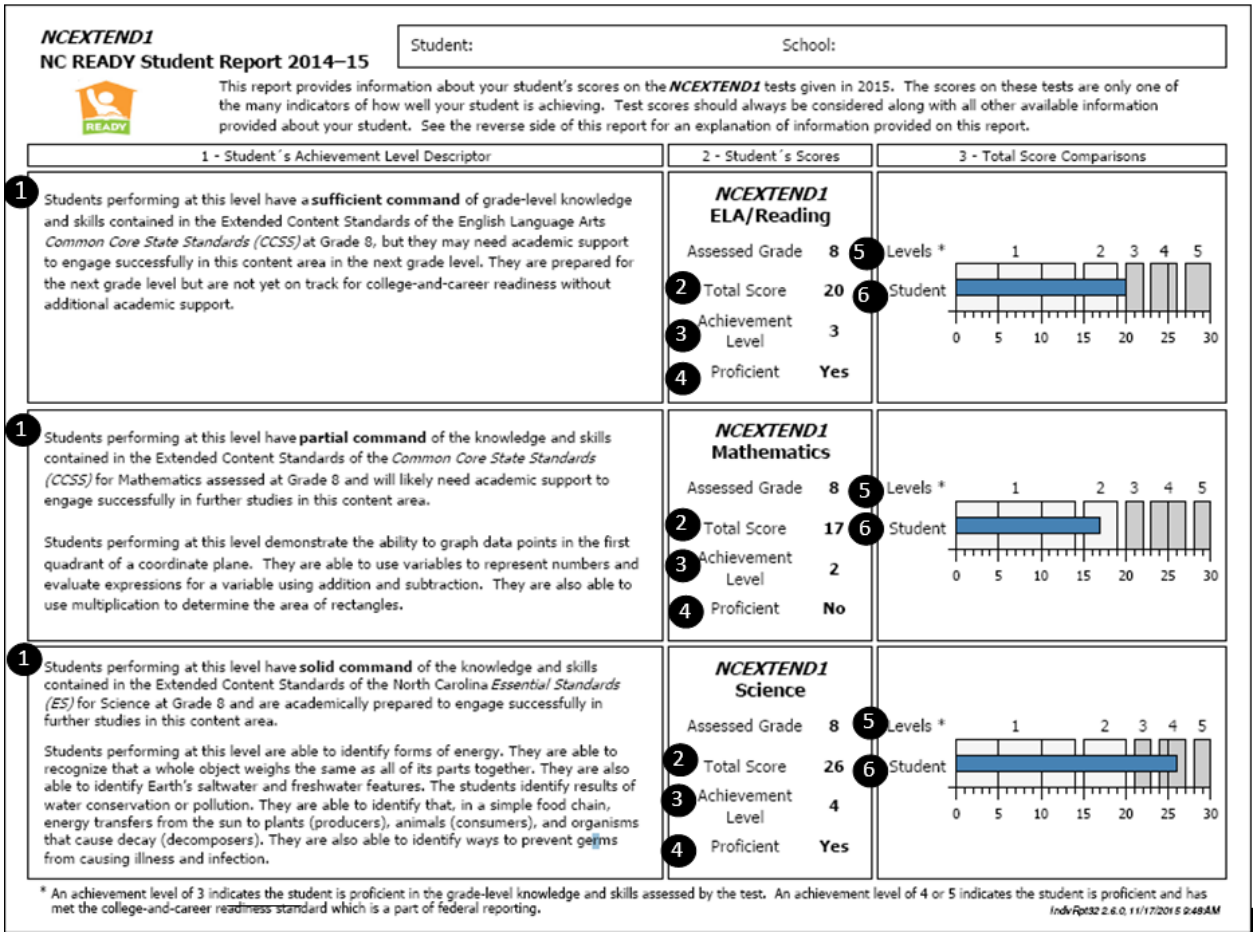
9.3 Sample Reports

To address fairness in reporting and ensure valid interpretation and use of individual test scores, NCDPI produces a series of custom reports along with interpretive guides. This ensures students, teachers, and stakeholders are able to make valid interpretations about test scores. The sample reports, along with the complete interpretive guide, is published on the NCDPI public webpage. This next section presents examples of the score reports with brief explanations of their use and interpretation.

9.3.1 Individual Student Reports (ISRs)

For students at grades 3, 4, 6, and 7, the Individual Student Report (ISR) for the **NCEXTEND1** *provides information concerning performance on the* **NCEXTEND1** *for ELA and Mathematics. For students at grades 5 and 8, the report includes information concerning performance on the* **NCEXTEND1***for ELA, Mathematics, and science. For students at grade 10, the report provides information for performance on the* **NCEXTEND1***English II, Math I and Biology. Example of ISR report is shown in Figure 9.18.* Key features are labeled and explained in the Index of Terms by Label Number section in the ISR.

Figure 9.18 Sample Individual Student Report for **NCEXTEND1** *Assessments*



The "Student's Achievement Level Descriptor" section (label 1) describes the level of achievement that the student is expected to have mastered given his or her assessment score. The achievement level descriptors can be viewed at http://www.ncpublicschools.org/accountability/testing//shared/achievelevel.

The Total Score (label 2) shows the number of questions the student answers correctly, also called a raw score. The Achievement Level (label 3) shows the level of achievement that the student is expected to have mastered given his or her assessment score. Five achievement levels (i.e., Levels 1, 2, 3, 4, and 5) are reported. Achievement Levels of 3 or above indicate grade-level proficiency (label 4). The Levels (label 5) refers to achievement levels, and label 6 represented by the blue horizontal bar shows a student's raw scores in relation to the achievement level cut score.

9.3.2 Class Roster Reports

The Class Roster Reports takes on many different combinations. A Class Roster Report can contain grade-specific student scores for each content area independently, or a class roster report can contain grade-specific student scores for combinations of content areas. The most typical combination for ***NCEXTEND1*** is a Class Roster Report that displays ELA, mathematics, and science scores together on one report for a specific grade. *Figure 9.19* displays a sample ***NCEXTEND1***Class Roster Report. This report is often produced at the class level and the school level and provides a summary report for students in the class.

Figure 9.19 Sample Class Roster Report for **NCEXTEND1**

PUBLIC SCHOOLS OF NORTH CAROLINA ALTERNATE ASSESSMENTS 2014-2015
NC Alternate Assessment NCEXTEND1 Class Roster
Regular test administration

LEASchCode = ??????????
DateTested = Mar 2015

SchoolName = ??????????

	Student Name	ELA/English II			Math I			Science/Biology		
		Assess Grade Level	Total Score (0-30) ②	Achieve Level ③	Assess Grade Level	Total Score (0-30) ②	Achieve Level ③	Assess Grade Level	Total Score (0-30) ②	Achieve Level ③
1	????? ??????	04	29	5	04	27	5			
2	???? ????????	05	30	5	05	30	5	05	30	5
3	??????? ?????	05	17	3	05	16	2	05	17	2
4	??????? ?????	03	24	4	03	25	4			
5	?????? ?????	05	29	5	05	28	5	05	27	5
6	?????????? ????????	03	30	5	03	27	4			
7	????????? ?????????	03	24	4	03	27	4			
8	??????? ?????????	03	16	2	03	16	2			
9	?????? ??????	03	30	5	03	30	5			
10	?????? ???????	04	30	5	04	30	5			
11	????? ????????????	03	14	1	03	16	2			
12	??????? ??????	05	21	4	05	25	4	05	26	5
13	?????????? ????	03	30	5	03	26	4			
14	?????? ?????	03	30	5	03	26	4			
15	??????? ?????	04	27	5	04	27	5			
16	?????? ????????	05	25	4	05	25	4	05	19	2
17	?????? ????????	03	30	5	03	26	4			
18	????????? ????????	04	27	5	04	28	5			
⑦	Group Mean		25.7			25.3			23.8	

English Language Arts/Reading and mathematics are assessed at grades 3-8.
Science is assessed at grades 5 and 8.
English II, Math I and Biology are assessed at grade 10.

The report is made up of three main summary columns:

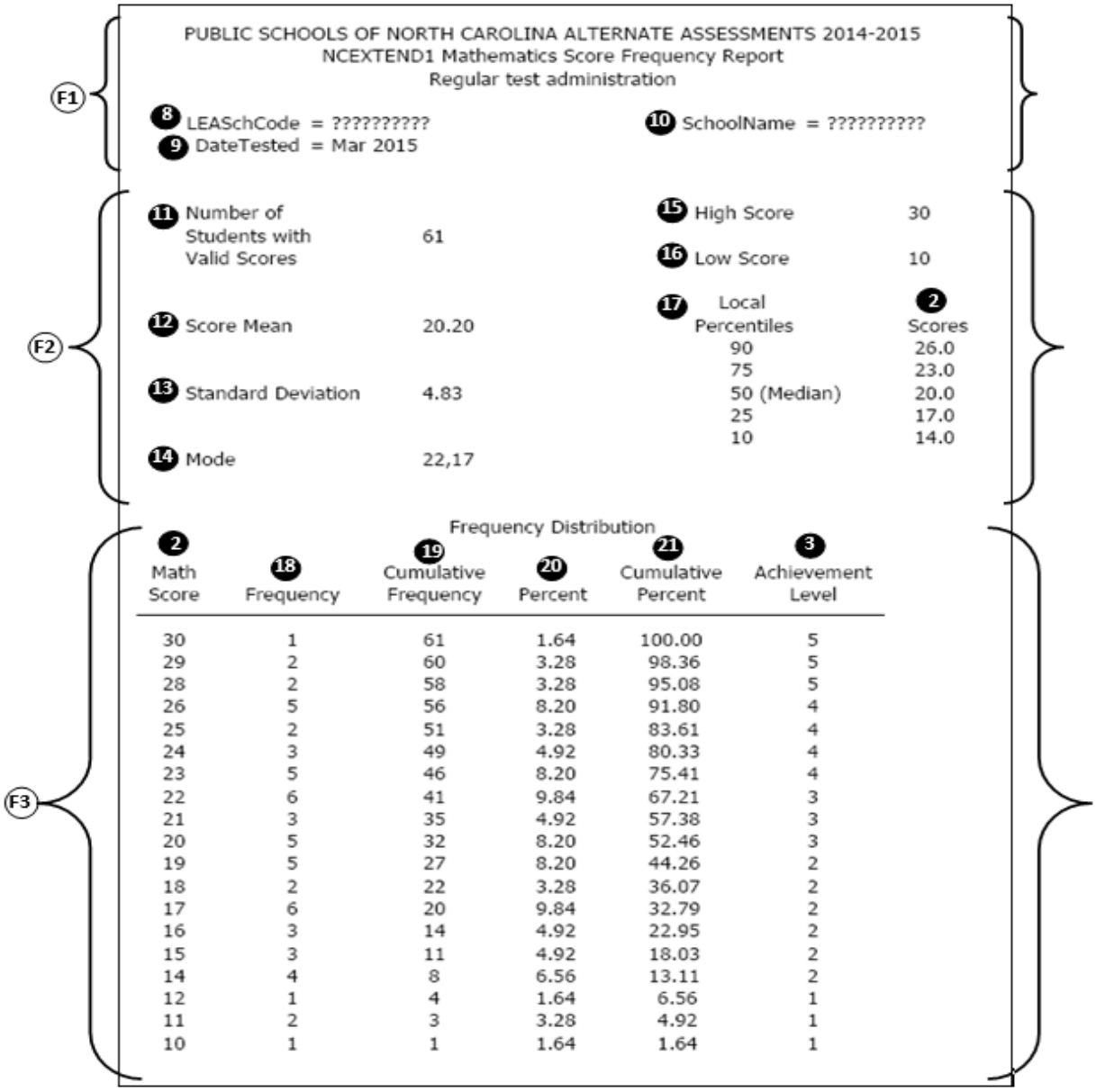
- Assess Grade Levels = shows the grade level for the ***NCEXTEND1***assessment
- Total Score = scores earned from the assessment
- Achieve Level = Reported achievement levels.

The Group Mean (label 7) shows the average of the group's scores. Group mean is the sum of all scores in the roster divided by the number of scores in the roster. For example, the group in the report got an averaged total score at 25.7 in ELA, 25.3 in math and 23.8 in science.

9.3.3 Score Frequency Reports

The Frequency Reports available in WinScan are often used to summarize score information at the class, school, system, and state levels. The WinScan Score Frequency Report presents the frequency, percent, cumulative frequency, and cumulative percent of each score at a specific grade. These reports can be created for each ***NCEXTEND1*** assessment. *Figure 9.20* presents a sample Score Frequency Report for an ***NCEXTEND1*** Mathematics Assessment. The ELA report is structured in a similar manner.

Figure 9.20 Sample Score Frequency Report for **NCEXTEND1** *Math*



PUBLIC SCHOOLS OF NORTH CAROLINA ALTERNATE ASSESSMENTS 2014-2015
NCEXTEND1 Mathematics Score Frequency Report
Regular test administration

F1

8 LEASchCode = ??????????
9 DateTested = Mar 2015
10 SchoolName = ??????????

F2

11 Number of Students with Valid Scores	61	15 High Score	30
		16 Low Score	10
12 Score Mean	20.20	17 Local Percentiles	2 Scores
		90	26.0
		75	23.0
13 Standard Deviation	4.83	50 (Median)	20.0
		25	17.0
		10	14.0
14 Mode	22,17		

F3

Frequency Distribution

2 Math Score	18 Frequency	19 Cumulative Frequency	20 Percent	21 Cumulative Percent	3 Achievement Level
30	1	61	1.64	100.00	5
29	2	60	3.28	98.36	5
28	2	58	3.28	95.08	5
26	5	56	8.20	91.80	4
25	2	51	3.28	83.61	4
24	3	49	4.92	80.33	4
23	5	46	8.20	75.41	4
22	6	41	9.84	67.21	3
21	3	35	4.92	57.38	3
20	5	32	8.20	52.46	3
19	5	27	8.20	44.26	2
18	2	22	3.28	36.07	2
17	6	20	9.84	32.79	2
16	3	14	4.92	22.95	2
15	3	11	4.92	18.03	2
14	4	8	6.56	13.11	2
12	1	4	1.64	6.56	1
11	2	3	3.28	4.92	1
10	1	1	1.64	1.64	1

The Score Frequency Report consists of three sections: the header (F1), a summary table for scores (F2), and a score frequency distribution table (F3).

The header specifies type of report (Score Frequency), the school year (2014–2015), and assessment type (***NCEXTEND1*** alternate assessment). The LEASchCode (label 8) indicates the Local Educational Agency school code; DateTested (label 9) refers to the time of year in which the exam was administered; and the SchoolName (label 10) indicates the school name.

The top row of the summary table (F2) indicates that 61 students in this report had valid scores (label 11). The highest score was 30 (label 15) and the lowest score was 10 (label 16). The mean score was 20.20 (label 12), the standard deviation was 4.83 (label 13), and there are two modes, 22 and 17 (label 14). The percentile scores are listed at the far right of the table (label 17). The scores are listed for the 10^{th}, 25^{th}, 50^{th}, 75^{th}, and 90^{th} percentiles (label 2). In this sample, a score of 23.0 corresponds to a percentile of 75. This means that 75 percent of the 61 students earned a score of 23.0 or less.

In the Frequency Distribution table (F3) the Math Score column (label 2) presents every score earned by the 61 students. The Frequency column (label 18) on the report presents the number of students that earned each score. For example, 4 students earned a score of 14. A "Missing" label would indicate that one student did not receive a score. The Cumulative Frequency column (label 19) displays the total number of students who earned up to and including a given score. This column shows 8 students earned up to and including a score of 14.

The Percent column (label 20) presents the percent of students that earned a given score (number of students that earned the score divided by total number of observations). This column shows that 6.56 percent of the students earned a score of 14. The Cumulative Percent column (label 21) displays the percent of students that earned up to and including a given score. This column shows 91.80 percent of the students earned up to and including a score of 26.

The Achievement Level column (label 3) displays the achievement level associated with each score. In this example, a score of 20 corresponds to an achievement level of 3.

Chapter 10 Validity Evidences and Reports 2012–2015

This chapter presents summary validity evidence collected in support of the interpretation of ***NCEXTEND1*** scores. The first couple of sections in this chapter present validity evidences in support of the internal structure of the ***NCEXTEND1*** assessments. Evidence presented in these sections includes reliability, standard error estimates, and classification consistency summary of reported achievement levels. The final sections of the chapter documents validity evidence based content summarized from the alignment study and external validity evidences based on relationship of scores with external variables.

10.1 Reliability Evidence of *NCEXTEND1*

Internal consistency reliability estimates provide a sample base summary statistic that describes the proportion of reported scores which is the true score variance. In order to justify valid use of scores in large scale standardized assessments, evidence must be documented that shows test results are stable, and replicable across all subgroups of the intended population. A reliable test produces scores that are expected to be relatively stable if the test were administered repeatedly under similar conditions. Scores from a reliable test reflect expected ability in the construct being measured with very little error variance. Internal consistency reliability coefficients (in this case measured by Cronbach's alpha) range from 0.0 to 1.0, where a coefficient of one refers to a perfectly reliable measures with no error. For large scale standardized high-stakes assessments, alpha estimates of 0.85 or higher are generally desirable.

Table 10.1 shows reliability estimates (Cronbach alpha) for different subject areas by grade and major demographic variables. Across all tests, reliability estimates from the 2012–2013 population range from 0.78 to 0.89 for ELA, 0.64 to 0.76 for Math, and 0.76 to 0.84 for Science. Subgroups reliabilities are also consistent across forms and subgroups in the same range as the overall estimates. Exception to this general trend are recorded in subgroup (Black and Hispanic) reliabilities for some grades (e.g. ELA grade 3) in which the subgroup alpha deviates from the overall estimate by over 0.05.

Table 10.1 **NCEXTEND1** *Reliabilities by Population and Subgroup 2012–13*

Content	Grade	Gender		Ethnicity			All
		Female	Male	Black	Hispanic	White	
ELA	Grade 3	0.88	0.89	0.91	0.82	0.88	0.89
	Grade 4	0.81	0.81	0.79	0.79	0.83	0.81
	Grade 5	0.83	0.83	0.82	0.85	0.82	0.83
	Grade 6	0.80	0.78	0.80	0.80	0.79	0.79
	Grade 7	0.86	0.84	0.83	0.87	0.86	0.85
	Grade 8	0.76	0.79	0.78	0.79	0.78	0.78
English II	Grade 10	0.80	0.81	0.76	0.78	0.84	0.80
Math	Grade 3	0.75	0.77	0.81	0.63	0.75	0.76
	Grade 4	0.73	0.74	0.68	0.73	0.78	0.74
	Grade 5	0.75	0.75	0.74	0.78	0.75	0.75
	Grade 6	0.70	0.69	0.69	0.76	0.68	0.69
	Grade 7	0.73	0.75	0.72	0.81	0.74	0.74
	Grade 8	0.61	0.65	0.61	0.66	0.63	0.64
Math I	Grade 10	0.65	0.71	0.64	0.73	0.74	0.70
Science	Grade 5	0.81	0.81	0.81	0.82	0.81	0.81
	Grade 8	0.82	0.85	0.86	0.85	0.82	0.84
Biology	Grade 10	0.73	0.77	0.70	0.75	0.80	0.76

10.2 Standard Error of Measurement

The standard error of measurement (SEM) for the overall test is computed by $SEM = SD * \sqrt{(1-\alpha)}$. SEM addresses the accuracy of examinees' classifications. It allows for a probabilistic statement to be made about the amount of precision on student's reported score. For example, if a student scores 100 in a test with SEM of 2, then it can be stated with a 68% certainty (1 standard error) that the student score is accurate within plus or minus 2 points. In other words, a 68% confidence interval for a score of 15 is 13–17. If that student were to be retested, his or her score would be expected to be in the range of 13–17 about 68% of the time.

The overall standard error of measurement for the ***NCEXTEND1*** assessments are provided in

Table 7.10. In CTT it is assume that the rate of measurement error is the same along the entire score scale. For example, a student with a score of 10 in grade 3 ELA will have the same amount of error associated with his/her score as a student with a score of 25. SEM for ELA range from 2.47 (Grade 3) to 2.78 (Grade 10), for Math the range is 2.84 (Grade 5) to 2.98 (Grade 8), and for Science 2.51 (Grade 8) to 2.78 (Grade 10).

10.3 Evidence of Classification Consistency

The *No Child Left Behind Act* of 2001 (2002) and subsequent *Race to the Top Act* of 2009 (2009) emphasized the measurement of adequate yearly progress (AYP) with respect to percentage of students at or above performance standards set by states. With this emphasis on the achievement level classification, a psychometric interest could be how consistently and accurately assessment instruments can classify students into the achievement levels. The importance of classification consistency as a measure of the categorical decisions when the test is used repeatedly has been recognized in the Standard 2.16 of the Standards for Educational and Psychological Testing (AERA, APA, & NCME, 2014) which states, "*When a test or combination of measures is used to make categorical decisions, estimates should be provided of the percentage of examinees who would be classified in the same way on two applications of the procedure."* (p. 46).

The methodology used for estimating the reliability of achievement-level classification decisions, as described in Hanson and Brennan (1990) and Livingston and Lewis (1995), provides estimates of decision accuracy and classification consistency. Classification consistency refers to "the agreement between classifications based on two non-overlapping, equally difficult forms of the test," and decision accuracy refers to *"the extent to which the actual classifications of test takers (on the basis of their single-form scores) agree with those that would be made on the basis of their true scores, if their true scores could somehow be known"* (Livingston & Lewis, 1995, P. 178). That is, classification consistency refers to the agreement between two observed scores, while classification accuracy refers to the agreement between observed and true scores.

The analyses are implemented using the computer program BB-Class.[11] The program provides results for both the Hanson and Brennan (1990) and Livingston and Lewis (1995) procedures. The Hanson and Brennan (1990) procedures assume that a *"test consists of* n *equally weighted, dichotomously-scored items"* while the Livingston and Lewis (1995) procedures are intended to handle situations where *"(a) items are not equally weighted and/or (b) some or all of the items are polytomously scored"* (Brennan, 2004, pp. 2–3), so the analyses for the ***NCEXTEND1*** tests followed the LL procedures.

Table 10.2 through *Table 10.4* present the decision accuracy and consistency indexes for achievement levels at each grade for ELA, math and science, respectively. Overall, the values indicate good classification accuracy (ranging from 0.83 to 0.98) and consistency (from 0.77 to 0.98). For example, if grade 3 ELA students who were classified as Level 2 take a non-overlapping, equally difficult form a second time, 89% of them would still be classified in Level 2.

Table 10.2 ELA **NCEXTEND1** *Classification Accuracy and Consistency Results*

ELA	Level 2 Partial Command		Level 3 Sufficient Command		Level 4 Solid Command		Level 5 Superior Command	
	Acc.	Con.	Acc.	Con.	Acc.	Con.	Acc.	Con.
Grade 3	0.92	0.89	0.90	0.85	0.89	0.85	0.88	0.84
Grade 4	0.95	0.92	0.89	0.84	0.88	0.83	0.88	0.84
Grade 5	0.94	0.91	0.91	0.87	0.88	0.84	0.88	0.83
Grade 6	0.93	0.90	0.89	0.84	0.87	0.82	0.88	0.87
Grade 7	0.93	0.89	0.88	0.83	0.88	0.83	0.89	0.85
Grade 8	0.94	0.91	0.89	0.84	0.86	0.81	0.88	0.87
English II	0.93	0.90	0.89	0.85	0.88	0.83	0.90	0.86

Note: Acc. = Classification Accuracy; Con. = Classification Consistency

[11] BB-Class is an ANSI C computer program that uses the beta-binomial model (and its extensions) for estimating classification consistency and accuracy. It can be downloaded from https://www.education.uiowa.edu/centers/casma/computer-programs#de748e48-f88c-6551-b2b8-ff00000648cd.

Table 10.3 Math **NCEXTEND1** *Classification Accuracy and Consistency Results*

Math	Level 2 Partial Command		Level 3 Sufficient Command		Level 4 Solid Command		Level 5 Superior Command	
	Acc.	Con.	Acc.	Con.	Acc.	Con.	Acc.	Con.
Grade 3	0.95	0.92	0.88	0.83	0.87	0.81	0.94	0.93
Grade 4	0.93	0.90	0.89	0.84	0.88	0.83	0.92	0.90
Grade 5	0.94	0.91	0.88	0.83	0.87	0.82	0.91	0.90
Grade 6	0.95	0.92	0.87	0.82	0.83	0.77	0.97	0.97
Grade 7	0.92	0.89	0.88	0.83	0.90	0.87	0.97	0.97
Grade 8	0.92	0.88	0.87	0.82	0.89	0.86	0.98	0.98
Math I	0.93	0.90	0.89	0.85	0.87	0.82	0.90	0.89

Note: Acc. = Classification Accuracy; Con. = Classification Consistency

Table 10.4 Science **NCEXTEND1** *Classification Accuracy and Consistency Results*

Science	Level 2 Partial Command		Level 3 Sufficient Command		Level 4 Solid Command		Level 5 Superior Command	
	Acc.	Con.	Acc.	Con.	Acc.	Con.	Acc.	Con.
Grade 5	0.96	0.93	0.89	0.85	0.87	0.82	0.86	0.82
Grade 8	0.95	0.93	0.89	0.85	0.87	0.82	0.84	0.81
Biology	0.96	0.93	0.90	0.86	0.87	0.81	0.84	0.81

Note: Acc. = Classification Accuracy; Con. = Classification Consistency

10.4 Alignment Study

The NCDPI commissioned the Wisconsin Center for Education Research (WCER) on September 2014 to conduct an in-depth study into the alignment of the ***NCEXTEND1*** assessments for mathematics, ELA and science to the state's Extended Content Standards for students with significantly cognitive disability as part of a larger effort to make a systemic examination of the state's standards-based reform efforts.

Specifically, this report focuses on describing the alignment characteristics of the ***NCEXTEND1*** assessment program in North Carolina based upon analyses of 17 assessment forms covering mathematics and ELA for grades 3, 4, 5, 6, 7, 8, and 10, as well as the ***NCEXTEND1*** science assessments for grades 5 and 8 and Biology grade 10.

10.4.1 Rationale for Standards Alignment

Standards-based educational reform has been the fundamental education model employed by states, and to a growing extent federal policymakers for twenty-plus years. Emerging out of the systemic research paradigm popular in the late eighties and early nineties, the standards-based model is essentially a systemic model influencing educational change. The standards-based system is based upon three fundamental propositions: 1) standards will serve as an explicit goal or target toward which curriculum planning, design, and implementation will move; 2) accountability for students, teachers and schools can be determined based upon student performance; and 3) standardized tests are aligned to the state content standards. Woven through these propositions is the notion of alignment, and the importance of it to the standards-based paradigm.

While examination of instructional alignment can help answer the first proposition, and alignment studies of assessments can help assure the third, neither of these approaches alone can address whether the assumptions of the second are justified. To do this, one must look at the role of both in explaining student achievement. Moreover, in order to address the overall effectiveness of the standards-based system as implemented in one or another location, one must be able to bring together compatible alignment indicators that span the domains of instruction, assessment, and student performance. The Surveys of Enacted Curriculum (SEC) is unique among alignment methodologies in that it allows one to examine the interrelationships of instruction, assessments, *and* student performance using an approach to examining alignment issues that is objective, systematic, low-inference, and quantifiable. The SEC, though best known for its tools for describing instructional practice, provides a methodology and set of data collection and analysis procedures that permit examination of all three propositions in order to consider the relationships between each. This allows for a look at the standards-based system as a whole to determine how well the system is functioning.

This document reports on Phase I of a three-phase study commissioned by the NCDPI to examine the effectiveness of the state's efforts to implement a newly structured standards-based system in the state. Phase I focuses on alignment of new assessments developed for mathematics and ELA in grades 3–8, as well as one high school end-of-course exam in each content area administered by the state. Phase II will focus on instructional alignment, and Phase III will examine student performance in light of students' opportunities to learn standards-based content given the assessments used to generate achievement results. Once all three phases have been completed, the state will have an in-depth look at its standards-based system, and it will have a wealth of information for considering its continuing efforts to provide quality educational opportunities to the state's K–12 population.

10.4.2 What Is Alignment Analysis?

Alignment, in terms of characteristics of assessment and instruction, is inherently a question about relationships. How does "A" relate to "B"? However, that also means alignment is inherently an abstraction in the sense that it is not easily measurable. As with most relationships, the answers to questions about alignment aren't ever as simple "yes" or "no," but rather they always contain a matter of degree. Relationships also tend to be multidimensional; they have more than a single aspect, dimension, or quality that is important for one to fully understand the nature of the alignment relationship. All of these factors make alignment analyses a challenging activity.

Alignment measures in SEC are derived from content descriptions. That is, alignment analyses report on the relationship between two multi-dimensional content descriptions. Each dimension of the two descriptions can then be compared, using procedures described below, to derive a set of alignment-indicator measures that summarizes the quantitative relationship between any two content descriptions on any of the dimensions used for describing academic content. In addition to allowing examination of each dimension independently, the following method allows for examination of alignment characteristics at the intersection of all three dimensions employed, producing a summative "overall" alignment indicator that has demonstrated a predictive capacity in explaining the variation of students' opportunities to learn assessed content, otherwise referred to as predictive validity.

Content descriptions appear in more detail in Section III. Note that two descriptions of academic content are collected in order to calculate and report alignment results: one a description of the content covered across a series of assessment forms for a particular grade level; and the other, a description of the relevant academic content standards for the assessed grade and subject. These content descriptions are systematically compared to determine the alignment characteristics existing between the two descriptions, using a simple iterative algorithm that generates an alignment measure or index based on the relevant dimension(s) of the content being considered.

As mentioned, there are three dimensions to the content descriptions collected, and hence three dimensions upon which to look at the degree of alignment the analyses indicate. These indicator measures can be distilled further to a single overall alignment index (OAI) that summarizes the alignment characteristics of any two content descriptions at the intersection of the three dimensions of content embedded in the SEC approach. These dimensions and the yielded alignment indicators are described next.

10.4.3 Alignment Targets

Typically, the alignment target for a state assessment is a set of academic content standards selected by the state. In some cases, the test framework or blueprint may be substituted as an alignment target, but this is not common as such documents often lack detailed content descriptions to support SEC style alignment analysis. States may decide to not assess some content areas based on logistical, economic or other factors. In such cases an assessment can look dramatically out of alignment due to the un-assessed content area(s) purposely omitted from the test. To better describe the alignment characteristics of such assessments there are some instances where a more targeted selection of subject matter content is warranted.

Consider the case of reading assessments in the state. Reading encompasses only one portion of the complete set of content standards for ELAR. Yet the state has for example, explicitly chosen to not assess writing and other language skills as part of their reading assessment program. Even at the secondary level, where the state does include open-ended response items in its end of course assessments, those items are not scored for writing content. As a result, holding the reading assessment accountable to the writing content emphasized in ELAR standards would inevitably result in low alignment results relative to the larger scope of ELAR

content, providing a misleading sense of the alignment of the reading assessments to content associated with language arts reading skills.

In order to make the alignment measures for the reading assessment more appropriate given the intended focus of those assessments, the results reported for reading below are based on alignment across the following ELAR content areas represented in the SEC taxonomy: Vocabulary; Text and Print Features; Fluency; Comprehension; Critical Reasoning; Author's Craft; Language Study; and Listening & Viewing. Excluded from the alignment analysis were content associated with the areas of Phonemic Awareness, Phonics, Writing Processes, Elements of Presentation, Writing Applications and Speaking & Presenting

In mathematics the curriculum tends to be more stratified. That is basic mathematical operations and topics such as fractions, decimals and measurement units are typically handled at lower grades, while topics like algebra, geometry, and trigonometry tend to be focused at higher grade levels. As a result, even though there may clearly be mathematics content not taught at one or another grade level, such content is typically excluded from the mathematics standards for that grade level. Thus a modification of the alignment target is not generally needed for mathematics.

Nonetheless specific circumstances can and do arise in mathematics, that make a modification in the definition or description of the alignment target a reasonable consideration when it helps to highlight the impact of specific mathematics content on alignment. An example of this is provided below for grade 7 mathematics.

10.4.4 The Dimensions of Alignment

SEC content descriptions are collected at the intersection of three dimensions: (1) topic coverage (2) performance expectation and (3) relative emphasis. These parallel the three alignment indices that measure the relationship between the two descriptions on one or another of these three dimensions: (1) Topical Coverage (TC); (2) performance expectations (PE); and (3) balance of representation (BR).

When considered in combination with one another that is when all three dimensions are included in the alignment algorithm, a fourth summary measure of 'overall alignment' can be calculated. The procedure for calculating alignment is discussed further on in the report, as a discussion of what constitutes 'good' alignment using the SEC approach.

In short, each alignment indicator is expressed on a scale with a range of 0 to 1.0—with 1.0 representing identical content descriptions (perfect alignment) and 0 indicating no content in common between the two descriptions, or perfect misalignment. For reasons discussed further below, a threshold measure is set at 0.5 for each of the four summary indicator measures. Above the threshold alignment is considered to be at an acceptable level, and below is considered weak or questionable, indicating that a more detailed examination related to that indicator measure is warranted. This does not necessarily mean that the assessment is not appropriately aligned, only that an explanation for the relatively low result is prudent. It means more information is needed.

10.4.5 Content Analysis Workshop

Content descriptions used to generate visual displays for the ***NCEXTEND1*** ELA, mathematics, and science were collected using content analysis. All content analysis work was conducted using teams of content analysts (educators with K–12 content and teaching with special need students' expertise). As opposed to the content analysis process used for the general test where a one-half-day content analysis training workshop was conducted prior to the process (see: http://www.dpi.state.nc.us/docs/accountability/testing/technotes/alignreport15.pdf), the ***NCEXTEND1*** assessments followed remote data collection procedure. Remote procedure refers to the analysts of each content analysis team working remotely from one another, using web, phone and other electronic media (e.g. email) for team discussions. This approach can be more cost efficient when large numbers of analysts are not required and the materials for analysis can be securely disseminated among team members to the satisfaction of the state's assessment director. For these analyses, the testing materials were delivered by express courier to the analysis team leader and then distributed to analysis team members as relevant to their subject area focus.

Each team received an initial introduction to the task and materials, and provided the opportunity to discuss and become familiar with the testing materials and procedures used for testing. All team members were veteran analysts, and did not require training. Most team members had worked together previously on content analysis tasks. All team members had the relevant content expertise, and the majority of members had experience as curriculum specialists with one or another state education agency

The independent results for each team member's analysis were sent to the project lead for entry into the Surveys of Enacted Curriculum (SEC) data system. Once the results from each team member were received and entered for a given assessment, the team was informed that their analysis results were ready for review. Analysts were then able to access through a password-protected login the results for each team member for each assessment item for the assessments analyzed. Team members discussed items either by conference call or email, depending upon team members' schedules and preferences. Each analyst had the ability to add, subtract or revise their results based on the review and group discussion.

Once the reviews were completed, the data was collected and processed in preparation for alignment analysis just as done for the earlier alignment study for the general statewide assessments. From that point on, the analysis procedures followed were identical to those utilized in the prior study.

The alignment analyses of any two content descriptions are based on detailed comparisons of the descriptive results collected during the content analysis process. While alignment results are based on a straightforward computational procedure and provide precise measures of the relationship between the descriptions in the test and the standards. Simple visual comparison of two content maps is often sufficient to identify the key similarities and differences between any two descriptions. For example, a simple visual comparison of the two maps presented in *Figure 10.8* for grade 3 ELA ***NCEXTEND1*** test suggests that, while distinctions can be identified, both have a generally similar structure which suggests reasonably good alignment of the two descriptions.

10.4.6 Topic Coverage

Topic coverage (TC) is a directly measurable dimension and examines how well the assessment matches the relevant standards in terms of the topics covered by each. The algorithm used to calculate topical concurrence provides a summary of the extent to which topics in the assessment match the topics embedded in the relevant standards. *Table 10.5* provides the summary results for TC across the mathematics and reading assessments analyzed for this study.

Table 10.5 Topic Coverage Index by Grade

Subjects	Grade						
	3	4	5	6	7	8	10
ELA	0.53	0.75	0.71	0.79	0.72	0.72	0.67
Math	0.60	0.66	0.72	0.75	0.53	0.73	0.84
Science	N/A	N/A	0.63	N/A	N/A	0.71	0.72

For the most part, the results presented in Table 10.5 suggest good to strong alignment, particularly for the English language arts alignment target and mathematics.

10.4.7 Performance Expectations

The SEC taxonomies enable descriptions of academic content based on two dimensions ubiquitous to the field of learning: knowledge and skills. When referencing standards this is frequently summarized with the statement "what students should know and be able to do". The 'what students should know' part refers to topics, while the 'be able to do' reference expectations for student performance, or performance expectations for short. The SEC taxonomies enable the collection of content descriptions on both of these dimensions, and together form the alignment 'target' for both assessments and instruction.

Just as one can examine alignment with respect to topic coverage alone, it is possible to examine and compare the performance expectations embedded in the content descriptions of assessments and standards in a similar manner. This alignment indicator is referred to as performance expectations (PE), and is a direct measure based on the five categories of expectations for student performance employed by the SEC. While the labels vary slightly from subject to subject, the general pattern of expectations follows this general division: 1) Memorization/Recall, 2) Procedural Knowledge, 3) Conceptual Understanding, 4) Analysis, Conjecture and Proof, and 5) Synthesis, Integration and Novel Thinking.Table 10.6 reports the performance expectation alignment measures across the assessed grades for mathematics, reading, and science. As with topic coverage this dimension is expressed as an index with a range of 0.0 to 1.0, with 0.5 indicating acceptable alignment.

Table 10.6 Performance Expectations Index by Grade

Subjects	Grade						
	3	4	5	6	7	8	10
ELA	0.85	0.81	0.78	0.73	0.71	0.77	0.58
Math	0.85	0.88	0.55	0.81	0.91	0.84	0.88
Science	N/A	N/A	0.78	N/A	N/A	0.77	0.75

As can be seen from Table 10.6, all but one subject/grade easily surpass the threshold measure. All of the assessments report good, and most quite strong alignment results, with respect to performance expectations.

10.4.8 Balance of Representation

Of the three content dimensions on which alignment measures are based, two are directly measured and one is derived. Two of the content dimensions are based upon observer/analyst reports of the occurrence of one or another content description. The derived measure concerns 'how much' and is based on the number of reported occurrences for a specific description of content relative to the total number of reports making up the full content description. This yields a proportional measure, summing to 1.00. The SEC refers to this 'how much' dimension as 'balance of representation' (BR).

As a summary indicator, (BR) is calculated as the product of two values: the portion of the assessment that targets standards-based content, multiplied times the portion of standards-based content represented in the assessment. For example, if 90% of an assessment (i.e. 10% of the assessment covers content not explicitly referenced in the standards) covered 40% of the standards for a particular grade level (i.e. 60% of the content reflected in the standards was not reflected in the assessment), the BR measure would be 0.36. As with all the summary indicator measures, reported here, the 'threshold' for an acceptable degree of alignment is placed at 0.50 or higher. This example would thus reflect an alignment measure that would bear further review, given this criterion. For a fuller discussion of the rationale for the 0.5 measure, the reader is referred to Section II: Conducting & Interpreting Alignment Analysis of the September 2015 North Carolina Alignment Study Report.

The influence of BR runs through all of the alignment indices, since the relative emphasis of content is the value used in making comparisons between content descriptions. In a very real

sense the dimensions of topic and performance expectation provide the structure for looking at alignment, while the balance of representation provides the proportional values placed in that structure.

For assessments, relative emphasis is expressed in terms of the proportion of score points attributed to one or another topic and/or performance expectation. When talking about standards, relative emphasis refers to the number of times a particular topic and/or performance expectation is noted across all the strands of a standard presented for a given grade and subject.

Table 10.7 summarizes the balance of representation results for the assessed standards. With only a few exceptions (grade 4 math and grade 8 science) the BR summary measures tend to fall well below the 0.5 threshold.

Table 10.7 Balance of Representation Index by Grade

	Grade						
Subjects	3	4	5	6	7	8	10
ELA	0.28	0.36	0.30	0.37	0.31	0.44	0.44
Math	0.29	0.50	0.28	0.48	0.23	0.20	0.34
Science	N/A	N/A	0.43	N/A	N/A	0.53	0.27

The NCEXTEND1 alternate assessments face a particularly difficult challenge in achieving a strong BR measure, as the assessment for each test is limited to fifteen items. Given the shortness of the assessment instrument and the psychometric need for multiple items to support assertions of proficiency for any one topic, it is not surprising that these numbers are quite low compared to the general assessment instruments which incorporate about five times the number of test items per instrument. The balance of representation issue is addressed further under the discussion of findings below.

This one measure provides only one piece of the alignment picture and it tells only part of the story. The other indicators provide other perspectives for viewing alignment that help to provide a more detailed picture of the alignment relationship between the NCEXTEND1 alternate assessments and their standards.

10.4.9 Overall Alignment Results

While the SEC approach allows for reporting and consideration of the results in terms of each of these three characteristics of alignment, the most powerful alignment indicator results when content is considered, and alignment is measured at the intersection of all three dimensions. It is the most challenging or rigorous of the alignment indicators because for a bit of content to be considered aligned, it must match the target on all three characteristics or dimensions. The resulting overall alignment index (OAI) has a range of 0.00 to 1.00 with 0.50 or higher indicating an acceptable level of alignment (just like the other alignment indices). Overall alignment results are reported in Table 4. Only grade 3 language arts and grade 7 mathematics report OAI's below the 0.50 mark. Overall alignment results are reported in *Table 10.8*.

Table 10.8 Overall Alignment Index by Grade

Subjects	Grade						
	3	4	5	6	7	8	10
ELA	0.49	0.61	0.56	0.61	0.55	0.58	0.52
Math	0.52	0.57	0.54	0.71	0.47	0.53	0.81
Science	N/A	N/A	0.52	N/A	N/A	0.52	0.61

Even at that, the language arts result is just shy of the 0.5 mark at 0.49 and the grade 7 math result is not far below at 0.47. Nonetheless, following the typical procedures for alignment analyses, these subthreshold results are examined further to better understand the nature of the alignment issues causing the sub-threshold results. *Table 10.9* summarizes the results on all four of the alignment indicators for the two sub-threshold assessments with respect to OAI.

Table 10.9 Overall Alignment Index by Grade and Subject

Grade	OAI	BR	TC	PE
Grade 3 ELA	0.49	0.28	0.53	0.85
Grade 7 Math	0.47	0.23	0.53	0.87

Table 10.9 shows that in each case the key alignment issue centered on balance of representation. Note that the TC and PE measures for both assessments exceed the 0.5 threshold while the BR results are substantially below that threshold.

10.4.10 Summary and Conclusion of Alignment Analysis

This study collected content descriptions of ***NCEXTEND1*** assessments covering grades 3–8 for mathematics and reading, as well as science assessments for grades 5 and 8 and grade 10 Biology. The resulting content descriptions were then compared in terms of alignment to content descriptions of the extended standards.

The alignment analyses reported above reveal a largely well-aligned set of ***NCEXTEND1*** alternate assessments compared to the extended content standards. Once the English language arts/reading alignment target content is adjusted to remove content related to writing the alignment indicators for topic coverage and performance expectations for all grades show levels of alignment exceeding the 0.5 threshold. Only one English language arts/reading alternate assessment (grade 3) reported an overall alignment index (OAI) less than 0.5, but just barely below 0.49.

The alignment story is almost identical for mathematics, with all of the ***NCEXTEND1*** mathematics alternate assessments reporting topic coverage and performance expectation results above 0.5, and only one assessment below the threshold for OAI (grade 7 OAI = 0.47). Even then, as discussed above, if the grade 7 alignment target is modified to exclude probability content, the OAI measure for grade 7 mathematics increases to 0.62. Alternately, including probability content in future grade 7 NCEXTEND1 alternate assessment would have a similar impact on the alignment result.

Nearly all of the assessments had trouble meeting the 0.5 threshold for balance of representation. However, as discussed, with only a 15 item assessment it is very difficult to assess a sufficient range of content to reach the 0.5 threshold while maintaining test reliability and validity. Given these constraints the ***NCEXTEND1*** alternate assessments all reveal very satisfactory alignment results. The content analysis contour maps are visually displayed for the ***NCEXTEND1*** ELA in *Figure 10.1* through *Figure 10.7,* mathematics in *Figure 10.8* through *Figure 10.14,* and science in *Figure 10.15* through *Figure 10.17.* Note that the content description maps provided in the figures are displayed along three axes or dimensions: the Y-axis, represented by the list of ***NCEXTEND1*** ELA topics presented to the right of the image, the X-axis represented by the five categories of performance expectations running across the bottom of the image, and the Z-axis (displayed by contour lines and color bands), indicating the relative

emphasis for each intersection of topic and performance expectation. These three dimensions form the foundational structure for describing and analyzing content using the SEC approach. Academic content is described in terms of the interaction of topic and performance expectations. By measuring each occurrence of some element of content (topic by performance expectation), a measure of the relative emphasis of each content topic as it appears in the content description can be obtained.

For example, *Figure 10.1* indicates that the topics with the strongest emphasis in North Carolina's grade 3 ***NCEXTEND1*** ELA assessable standards (map to the right "Target Content Areas") are vocabulary, comprehension, critical reading, and language study particularly at the performance level of "Generate" and "Explain." A careful visual review of the two maps for ***NCEXTEND1*** grade 3 ELA shown in *Figure 10.1* in terms of the three alignment indices are described below. Similar interpretations can be followed for the other grades and content areas.

Topic Coverage (TC): topics with the strongest emphasis are vocabulary, comprehension, critical reading and Language Study, where the contour lines are closer together. This indicate the assessment blueprint is aligned to the content standards with respect to TC. The TC index for NCEXTEND1 grade 3 ELA is 0.53 above the threshold of 0.50 (see Table 10.5).

- Performance Expectation (PE): PE focuses on what students should "*be able to do*" more generally summarized by DOK levels. From the ***NCEXTEND1*** grade 3 ELA assessment map (left) the two strongest topics of emphasis are mostly assessed with "generate" and "explain" type items. The expectation of the standards focus is also on "generate" and "explain." Therefore, the performance expectation alignment is strong with PE of 0.85 (see *Table 10.6*).
- Balance of Representation (BR): The two figures are shaped similar (*Figure 10.1*). However, the contour of the target map is dense than the test map indicating that the balance of representation for ***NCEXTEND1*** grade 3 ELA assessment was low. This is also confirmed by a BR index of 0.28 (see *Table 10.7*).
- Overall Alignment Index (OAI): Only grade 3 language arts (0.49) and grade 7 mathematics (0.47) report OAI's (see *Table 10.8*) below the 0.50 mark indicating that overall alignment is reasonable given the fewer number of items in the test for the 1% population who are cognitively challenged.

Figure 10.1 Grade 3 **NCEXTEND1** *ELA Assessment and Standard Content Map*

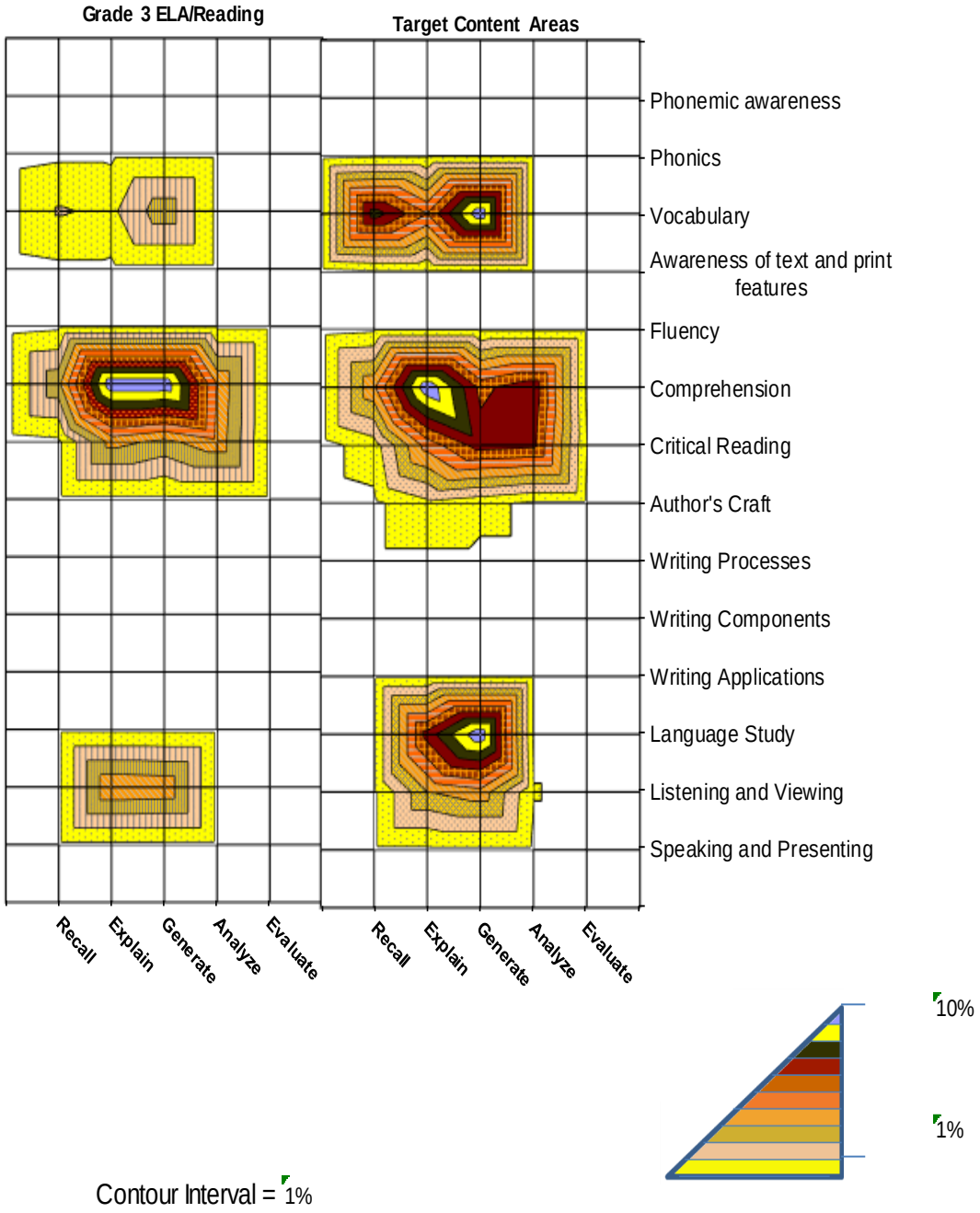


Figure 10.2 Grade 4 **NCEXTEND1** *ELA Assessment and Standard Content Map*

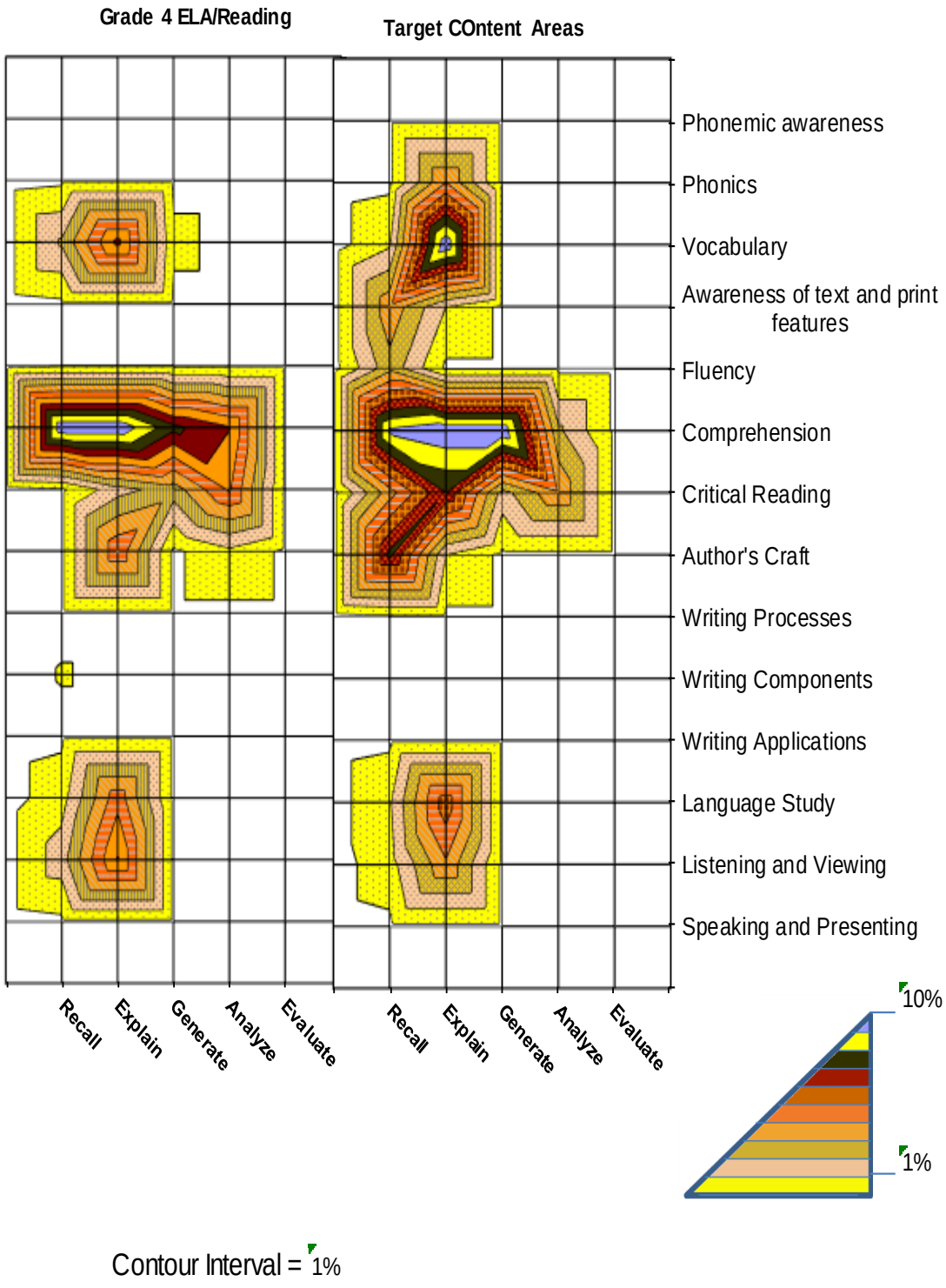


Figure 10.3 Grade 5 **NCEXTEND1** *ELA Assessment and Standard Content Map*

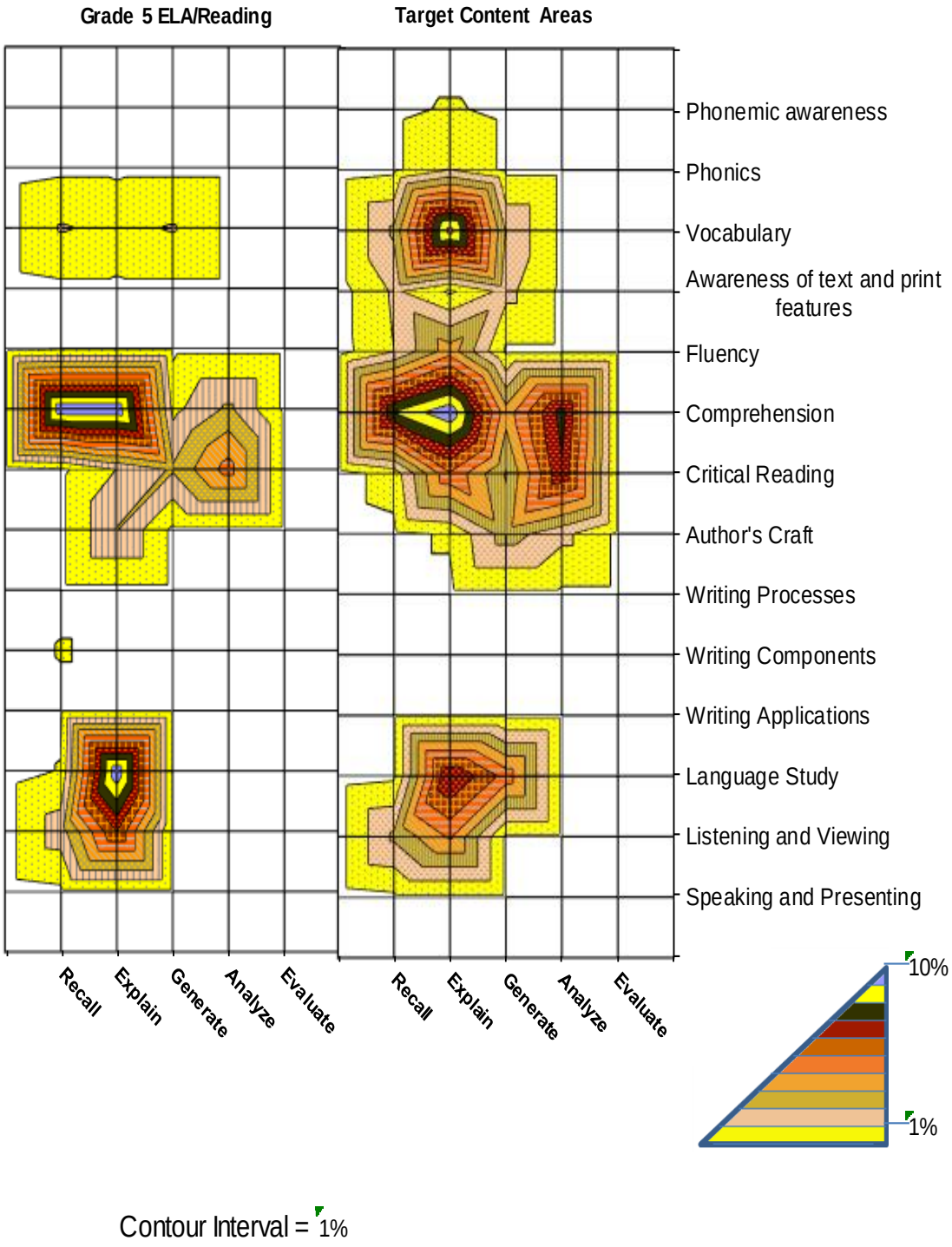


Figure 10.4 *Grade 6* **NCEXTEND**1 *ELA Assessment and Standard Content Map*

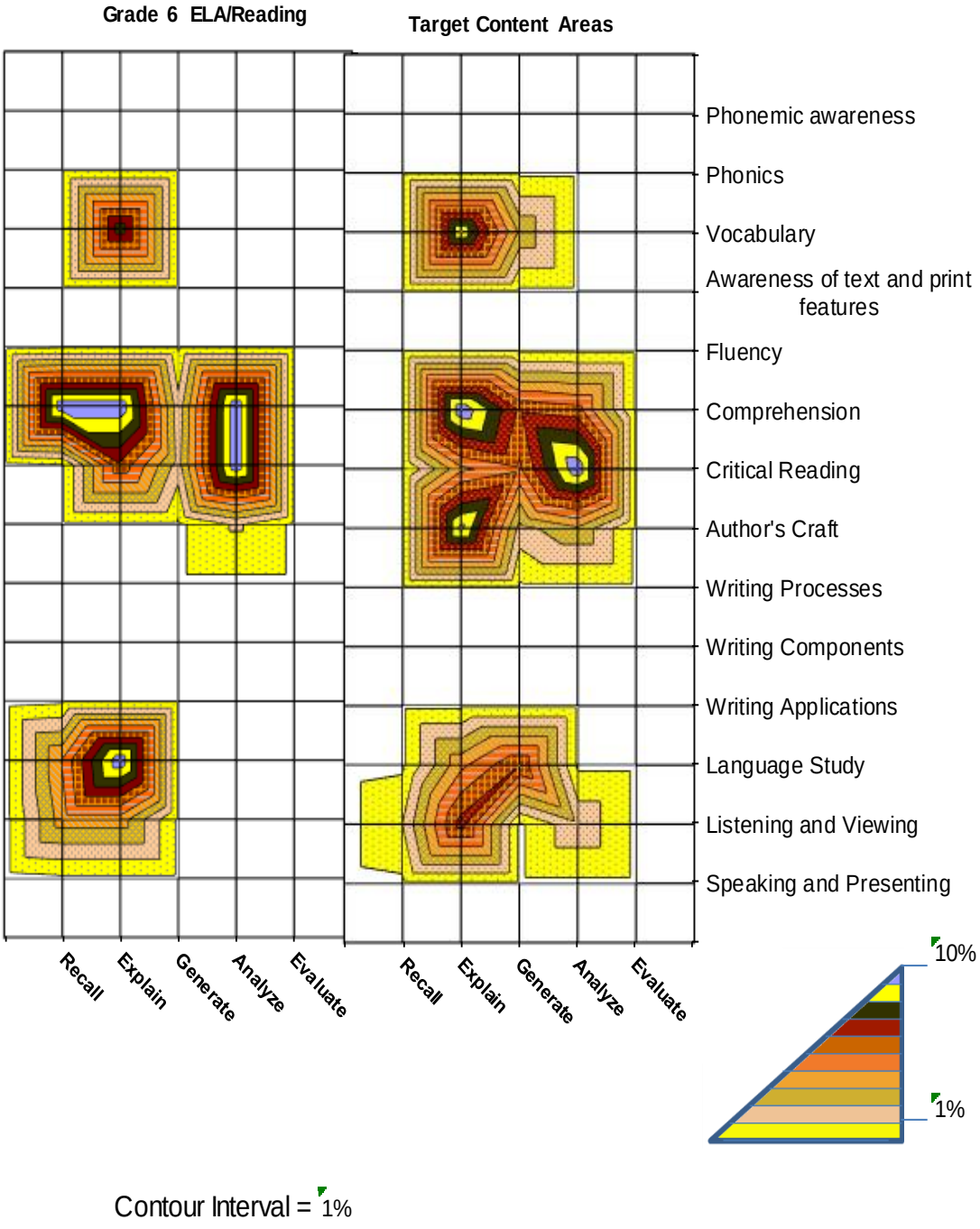


Figure 10.5 Grade 7 **NCEXTEND1** *ELA Assessment and Standard Content Map*

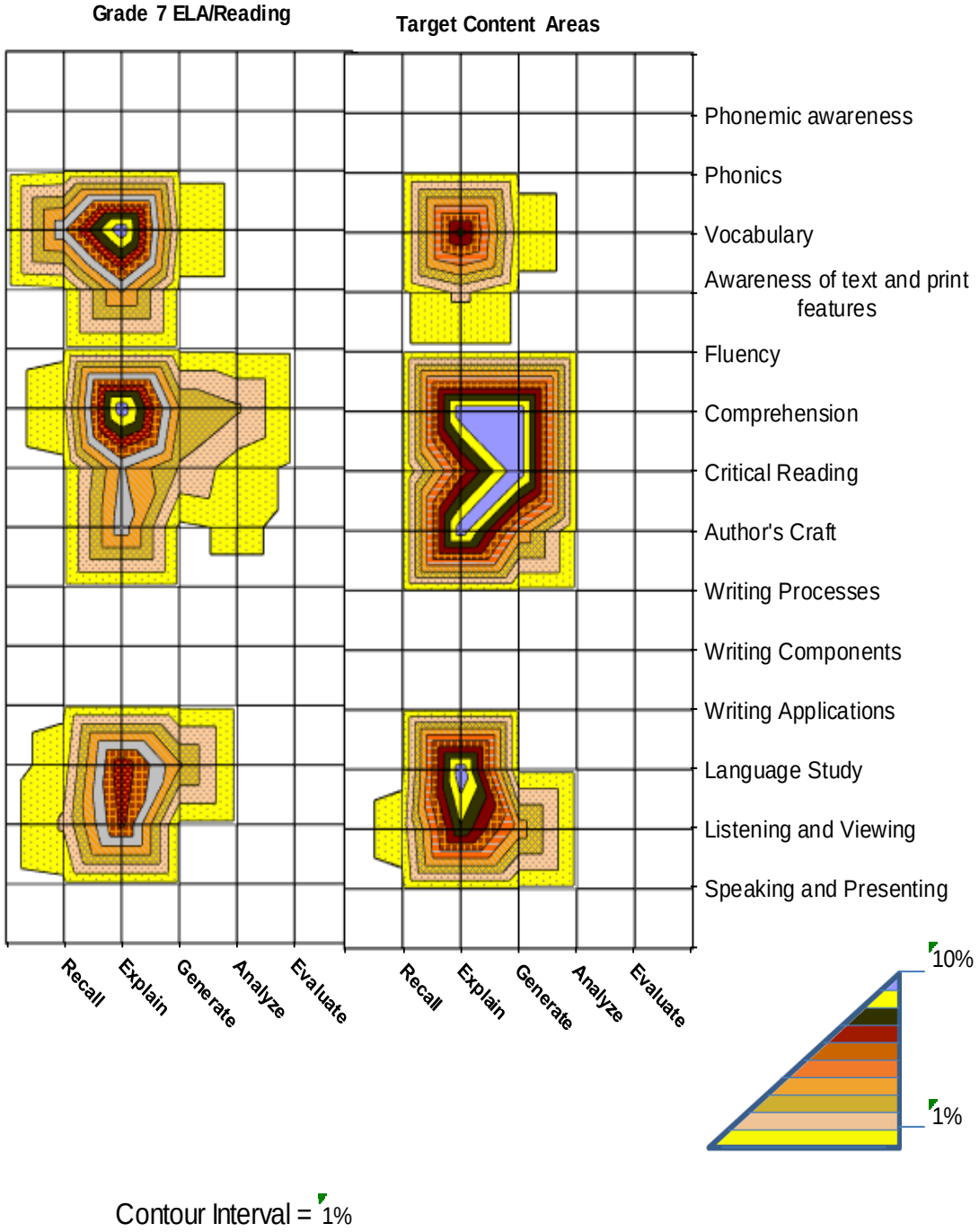


Figure 10.6 Grade 8 **NCEXTEND1** *ELA Assessment and Standard Content Map*

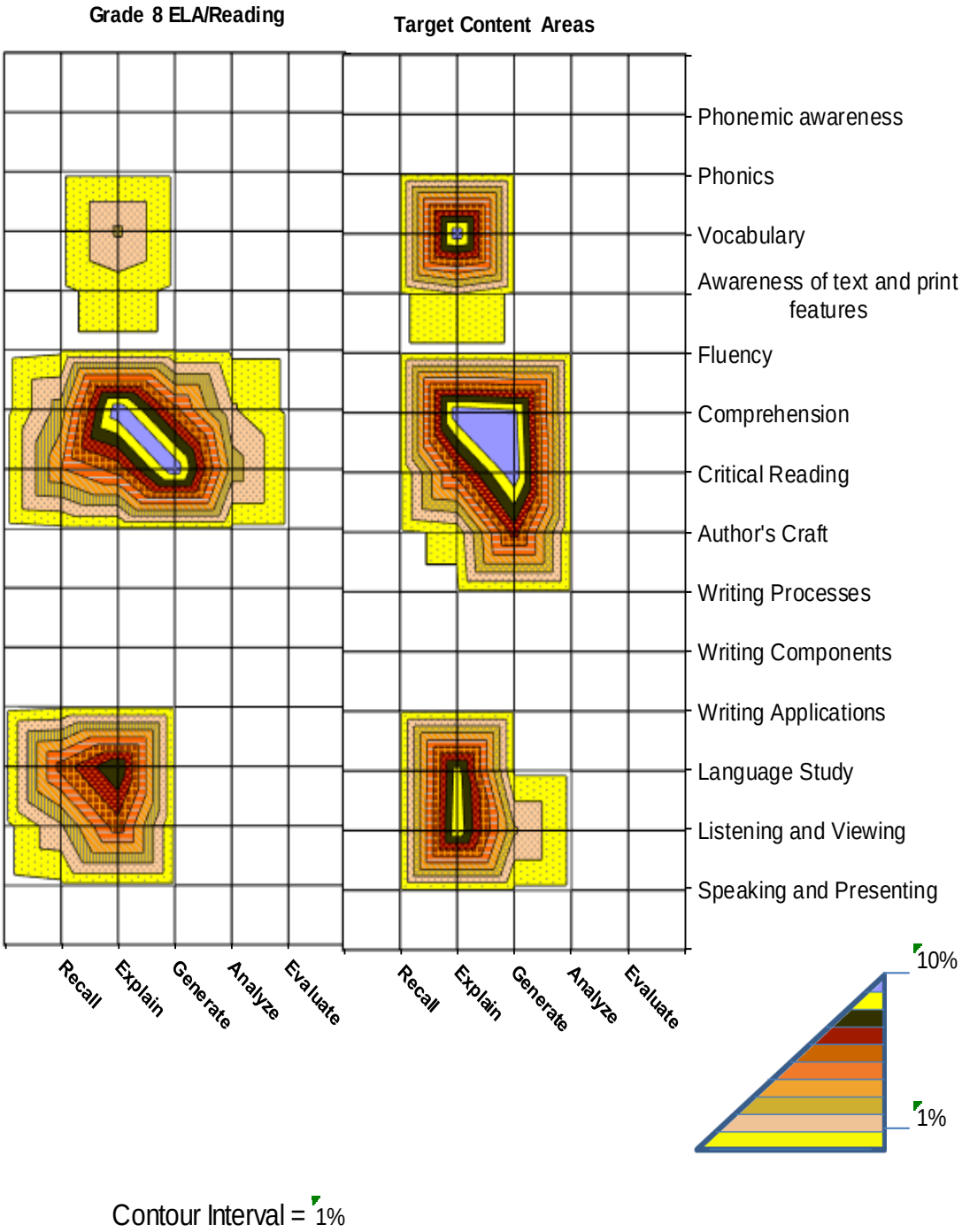


Figure 10.7 Grade 10 **NCEXTEND1** *English II Assessment and Standard Content Map*

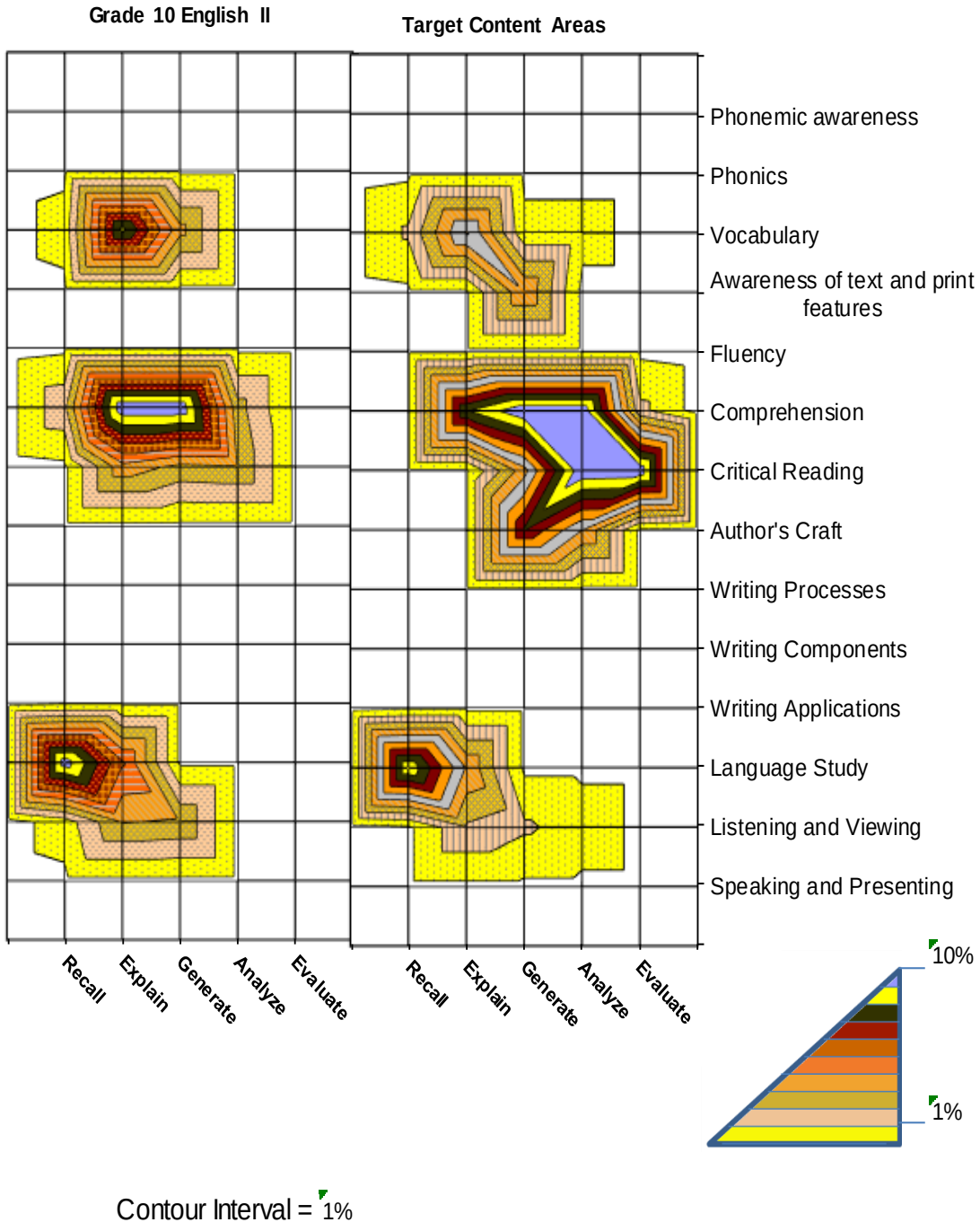


Figure 10.8 Grade 3 **NCEXTEND1** *Math Assessment and Standard Content Map*

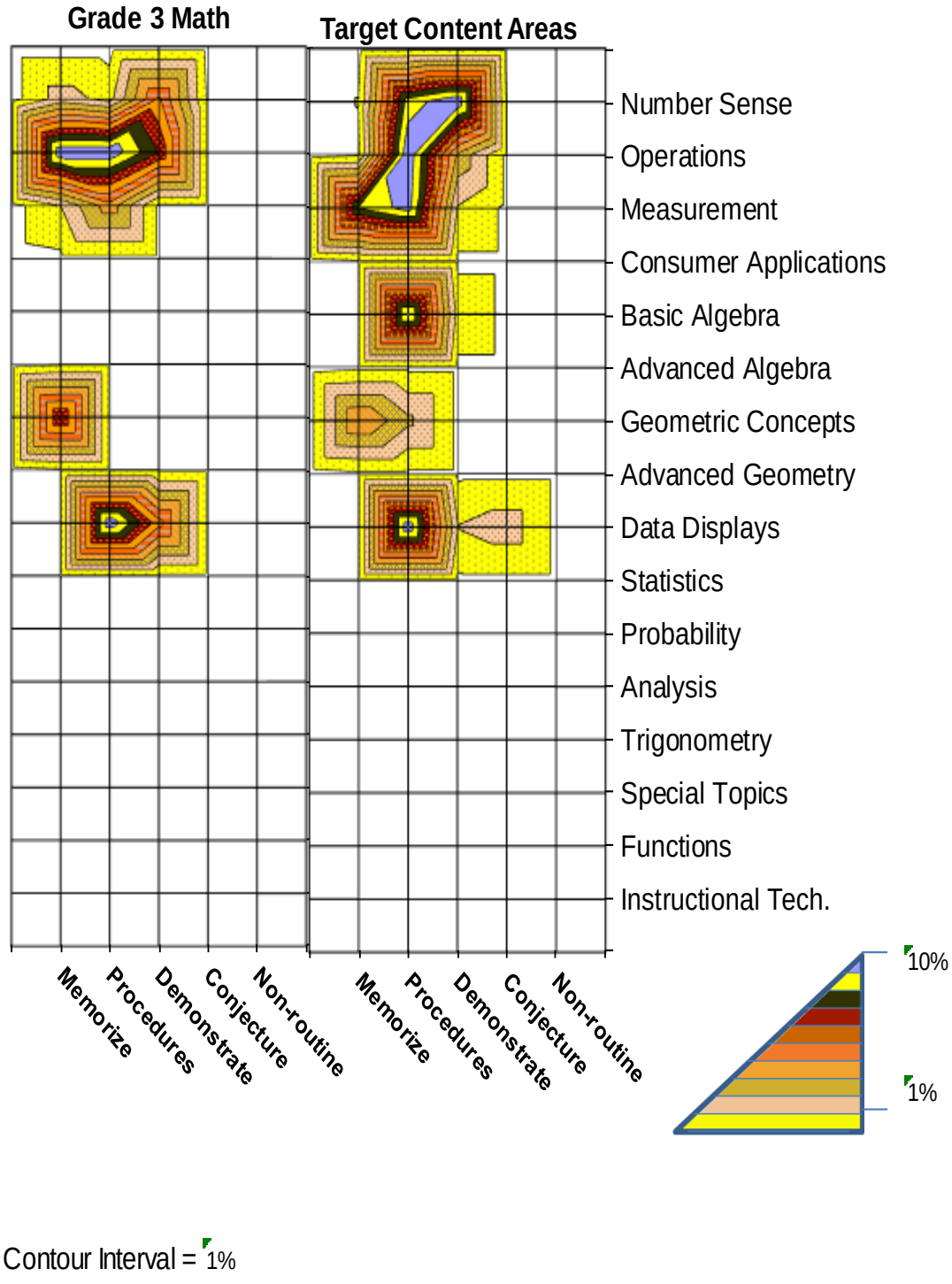


Figure 10.9 Grade 4 **NCEXTEND1** *Math Assessment and Standard Content Map*

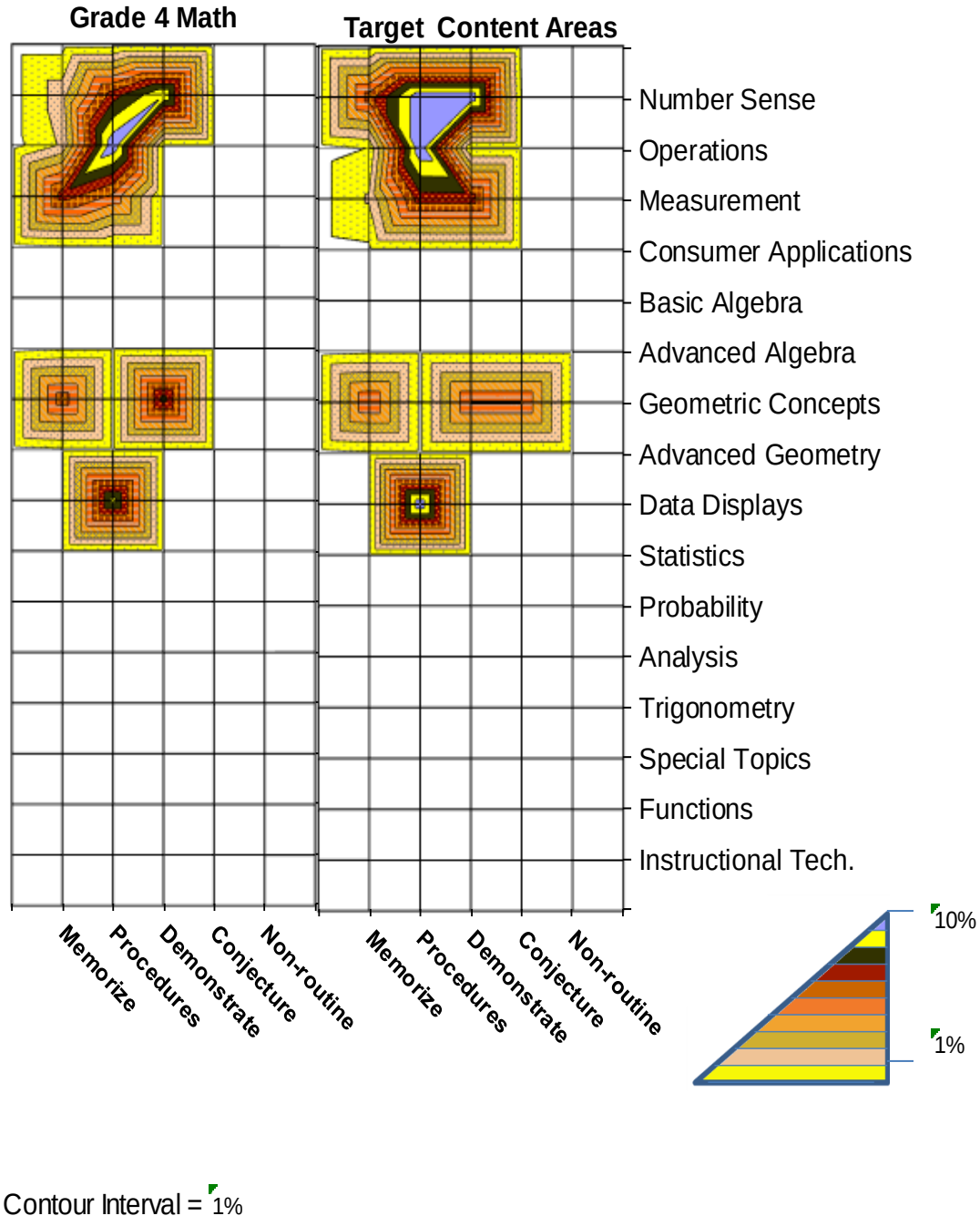


Figure 10.10 Grade 5 **NCEXTEND1** *Math Assessment and Standard Content Map*

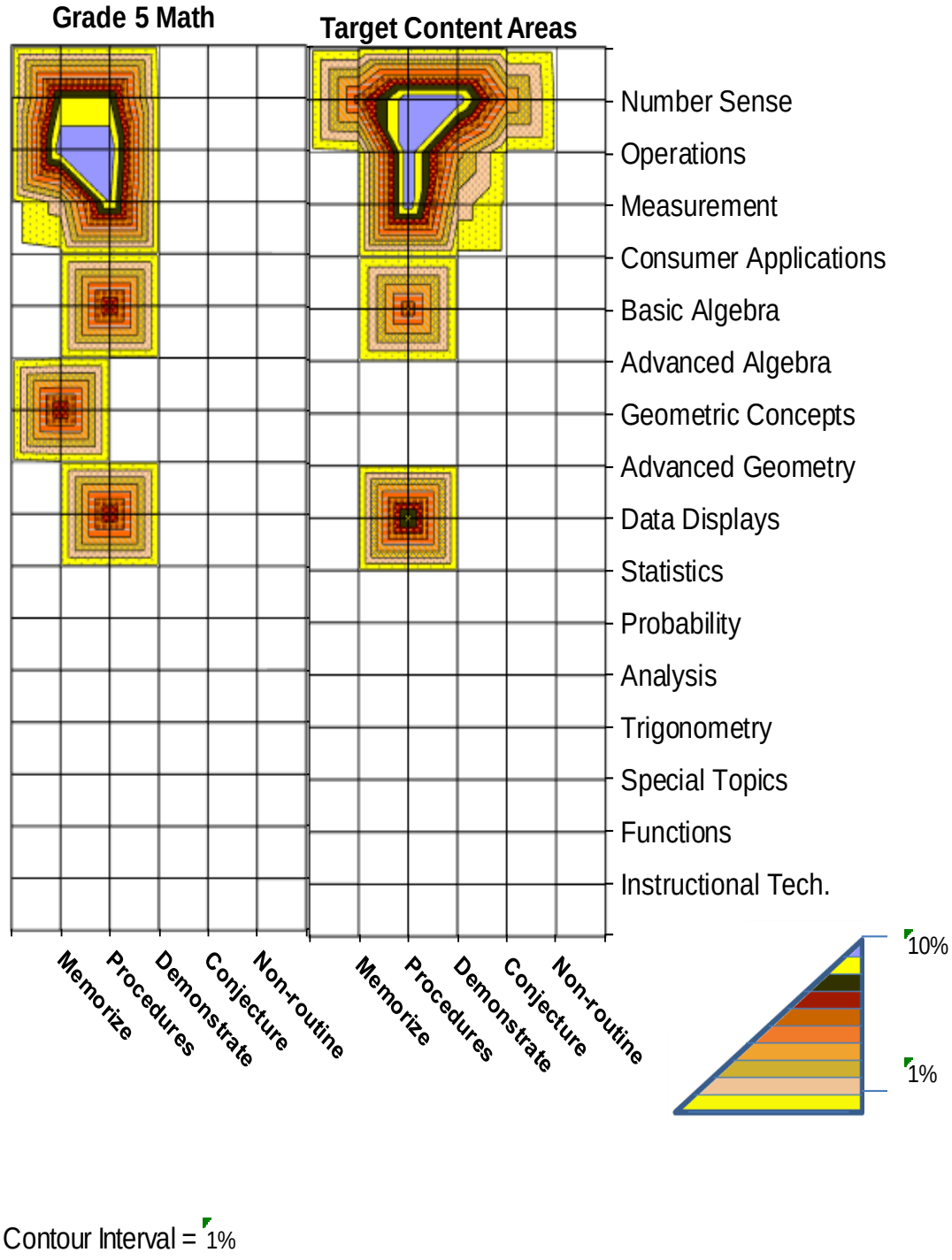


Figure 10.11 Grade 6 **NCEXTEND1** *Math Assessment and Standard Content Map*

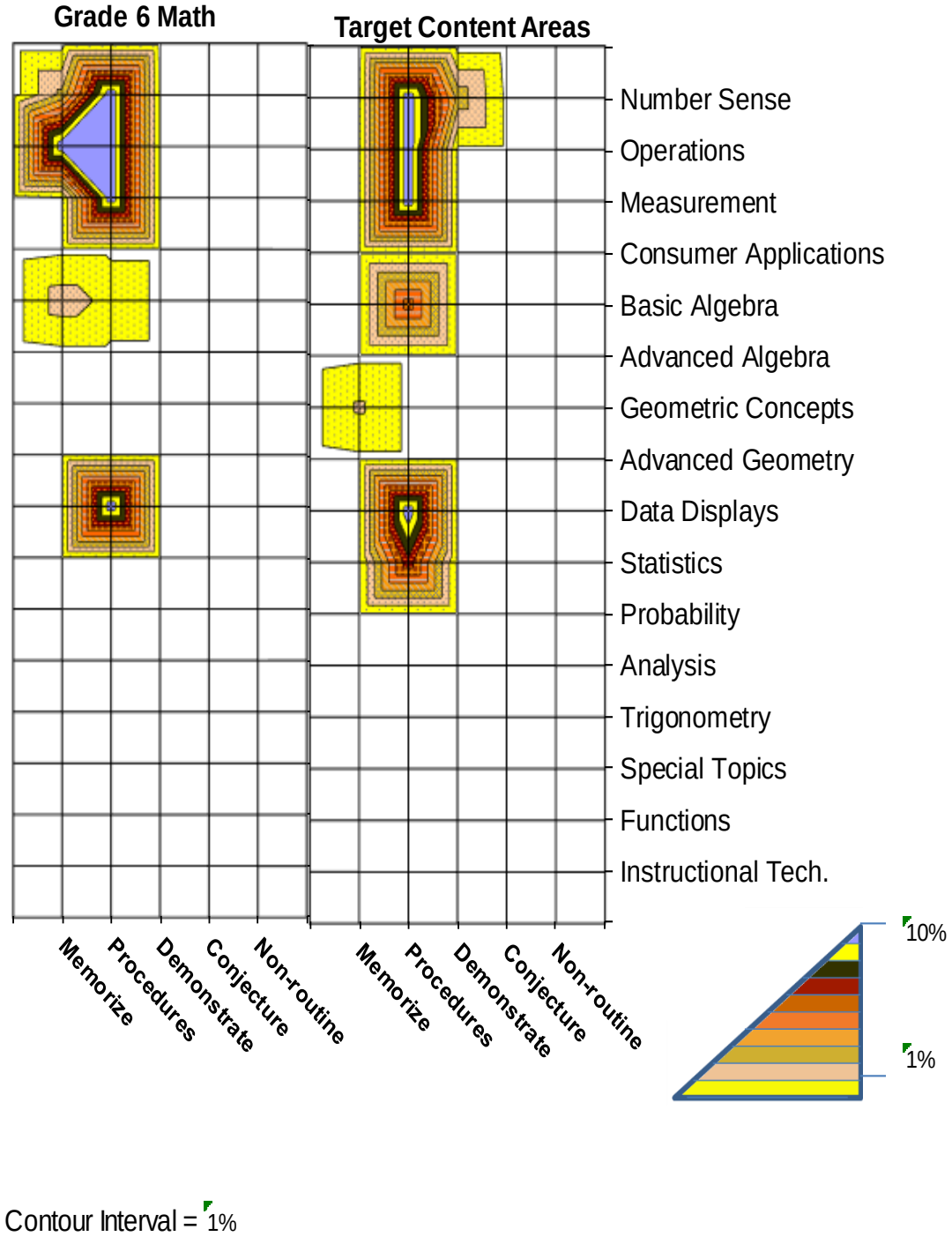


Figure 10.12 Grade 7 **NCEXTEND1** *Math Assessment and Standard Content Map*

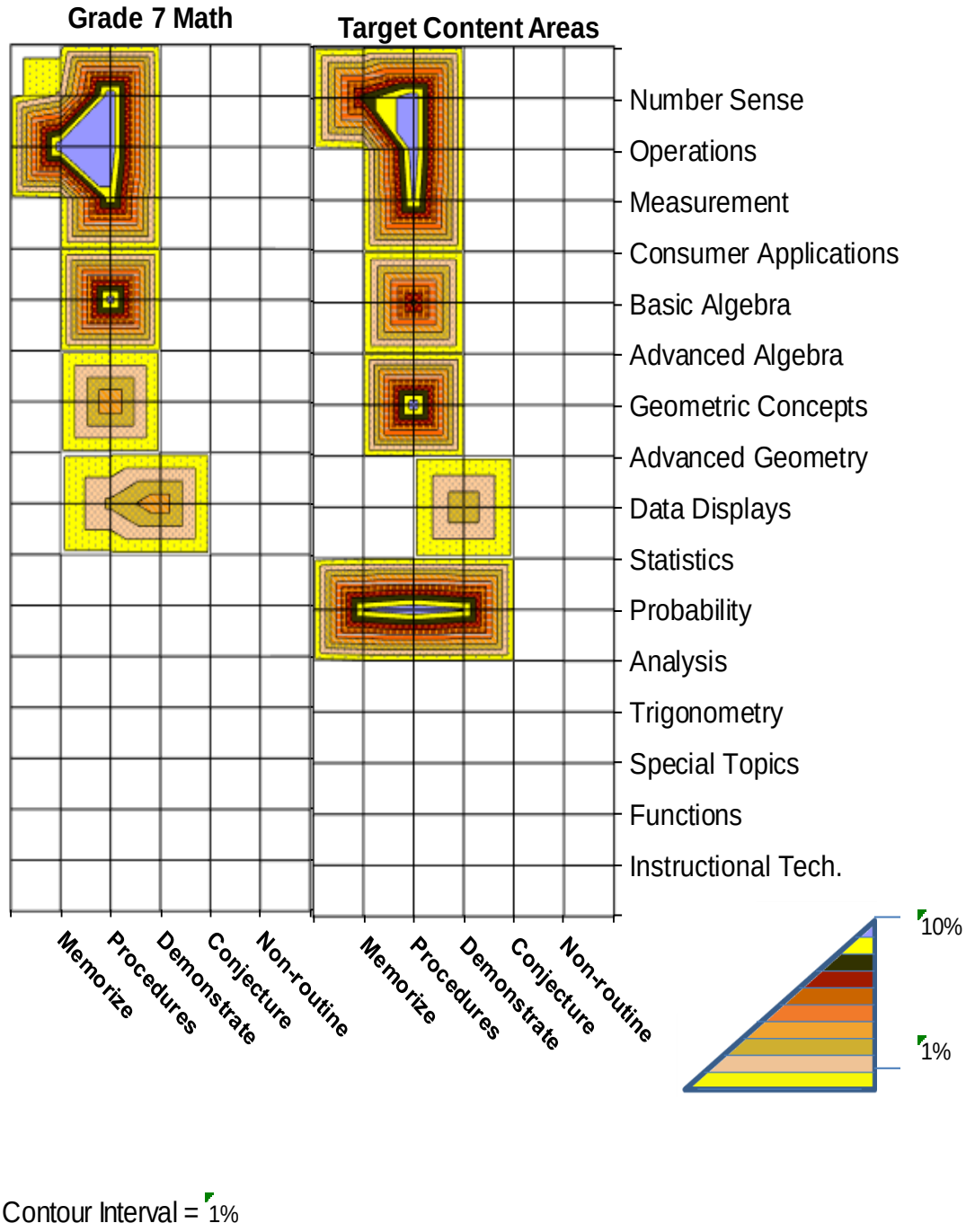


Figure 10.13 Grade 8 **NCEXTEND1** *Math Assessment and Standard Content Map*

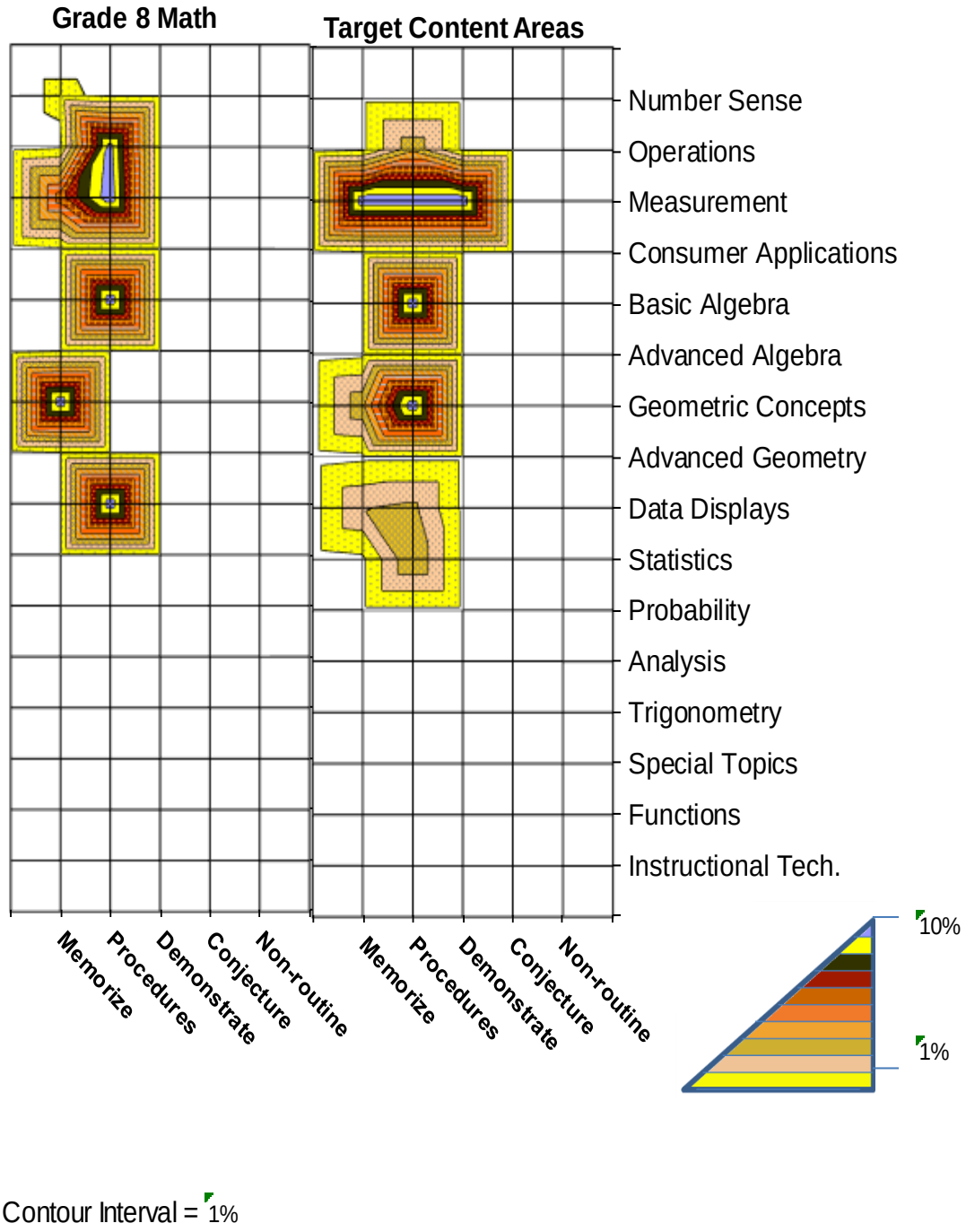


Figure 10.14 Grade 10 **NCEXTEND1** *Math I Assessment and Standard Content Map*

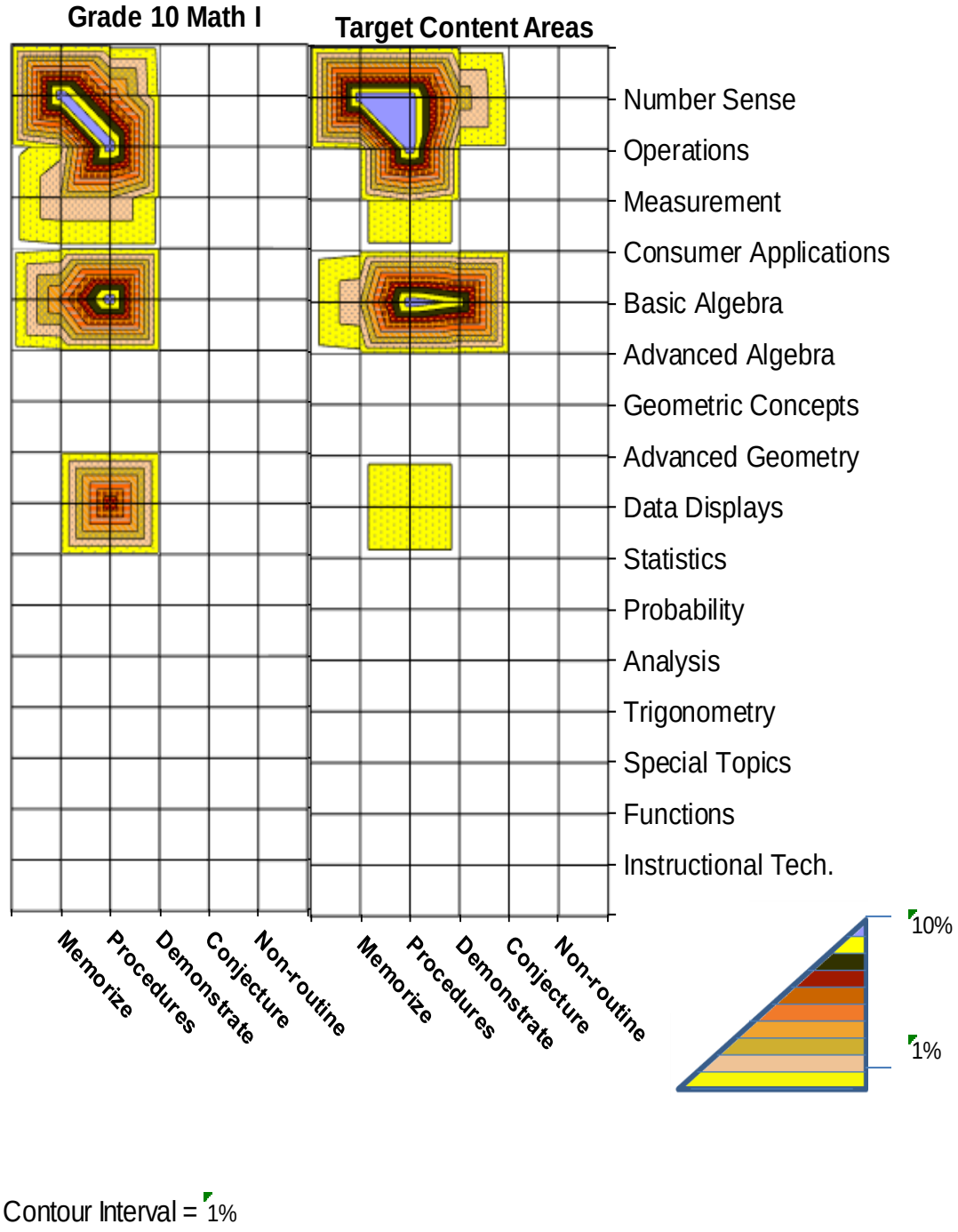
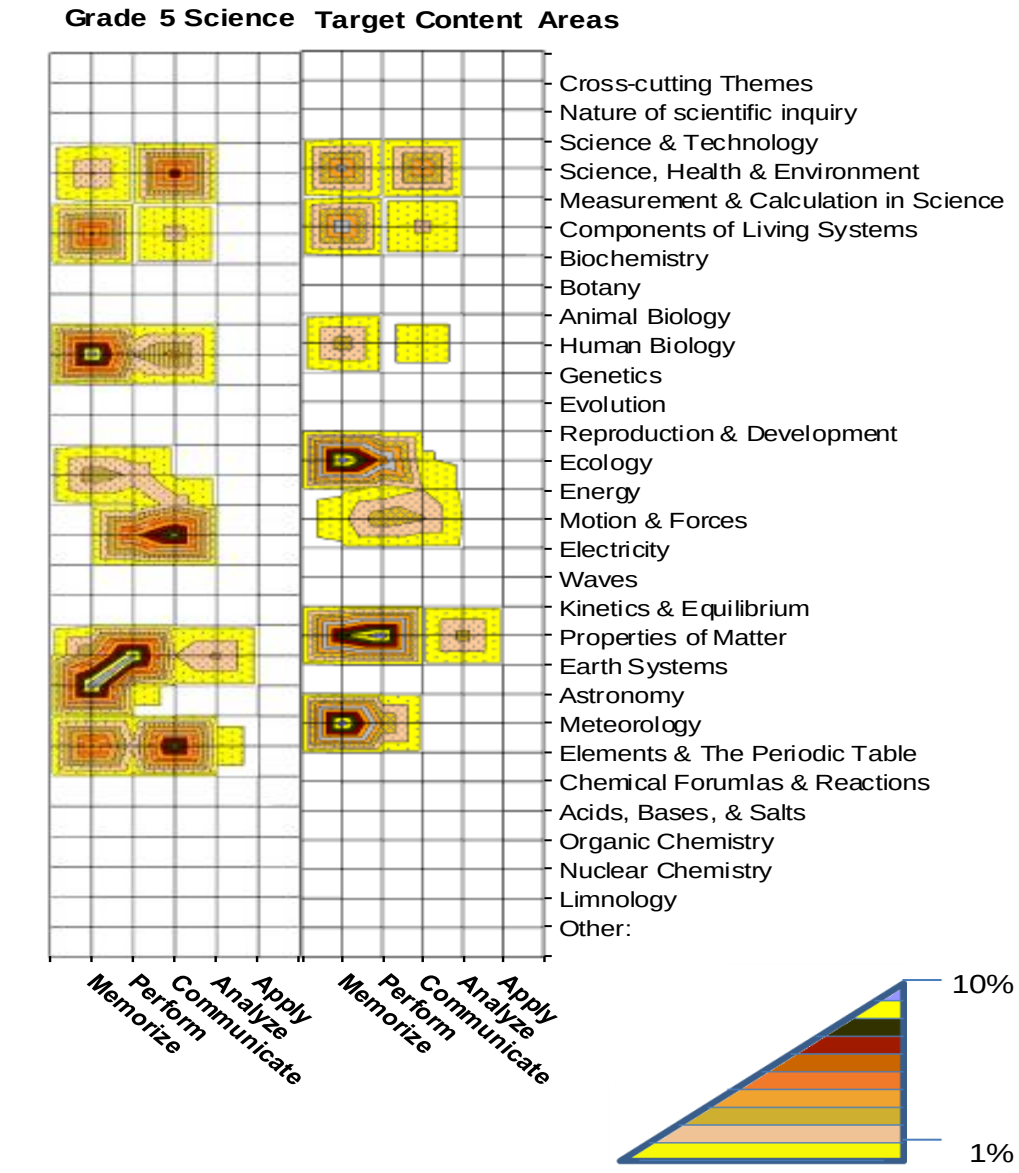
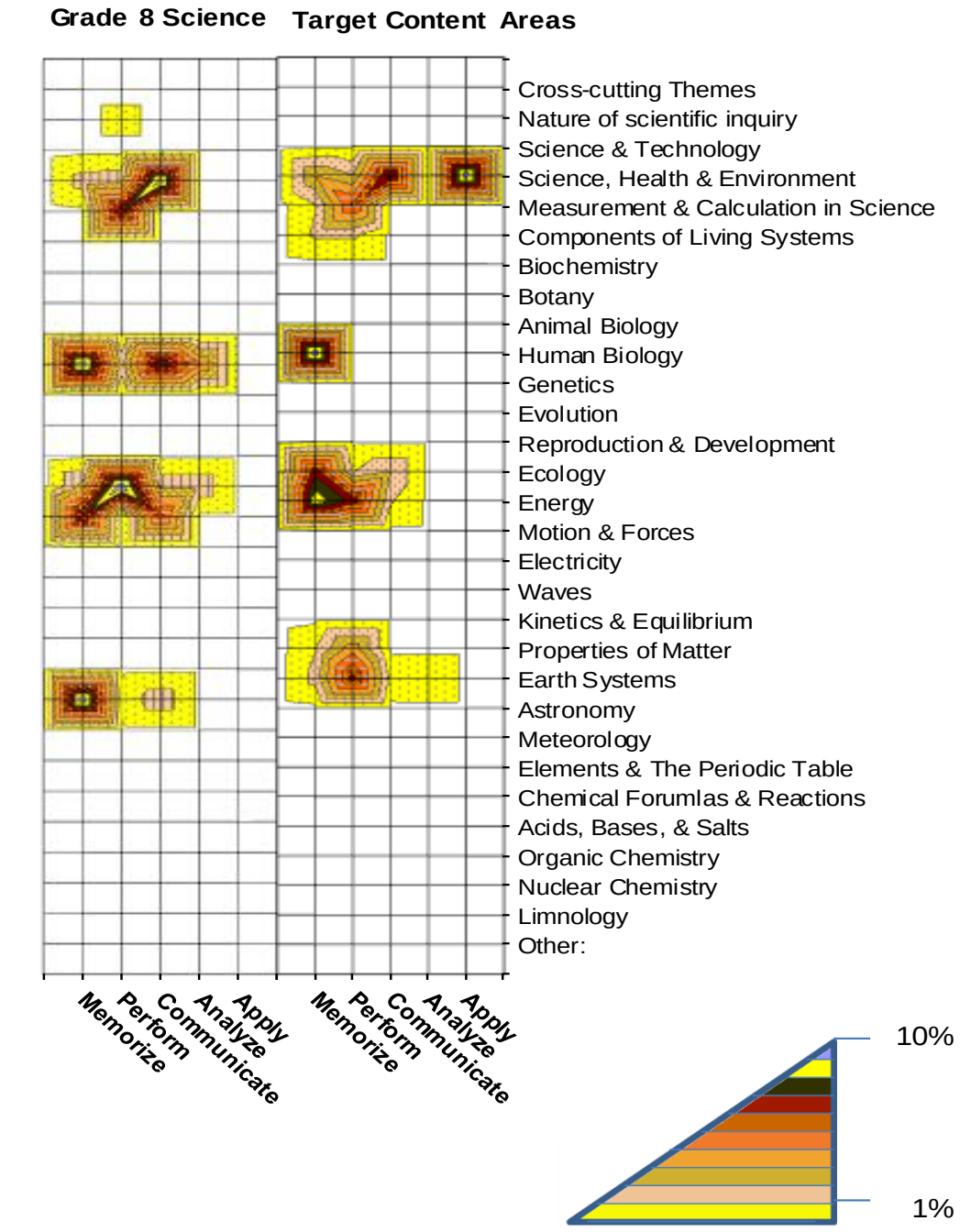


Figure 10.15 Grade 5 **NCEXTEND1** *Science Assessment and Standard Content Map*



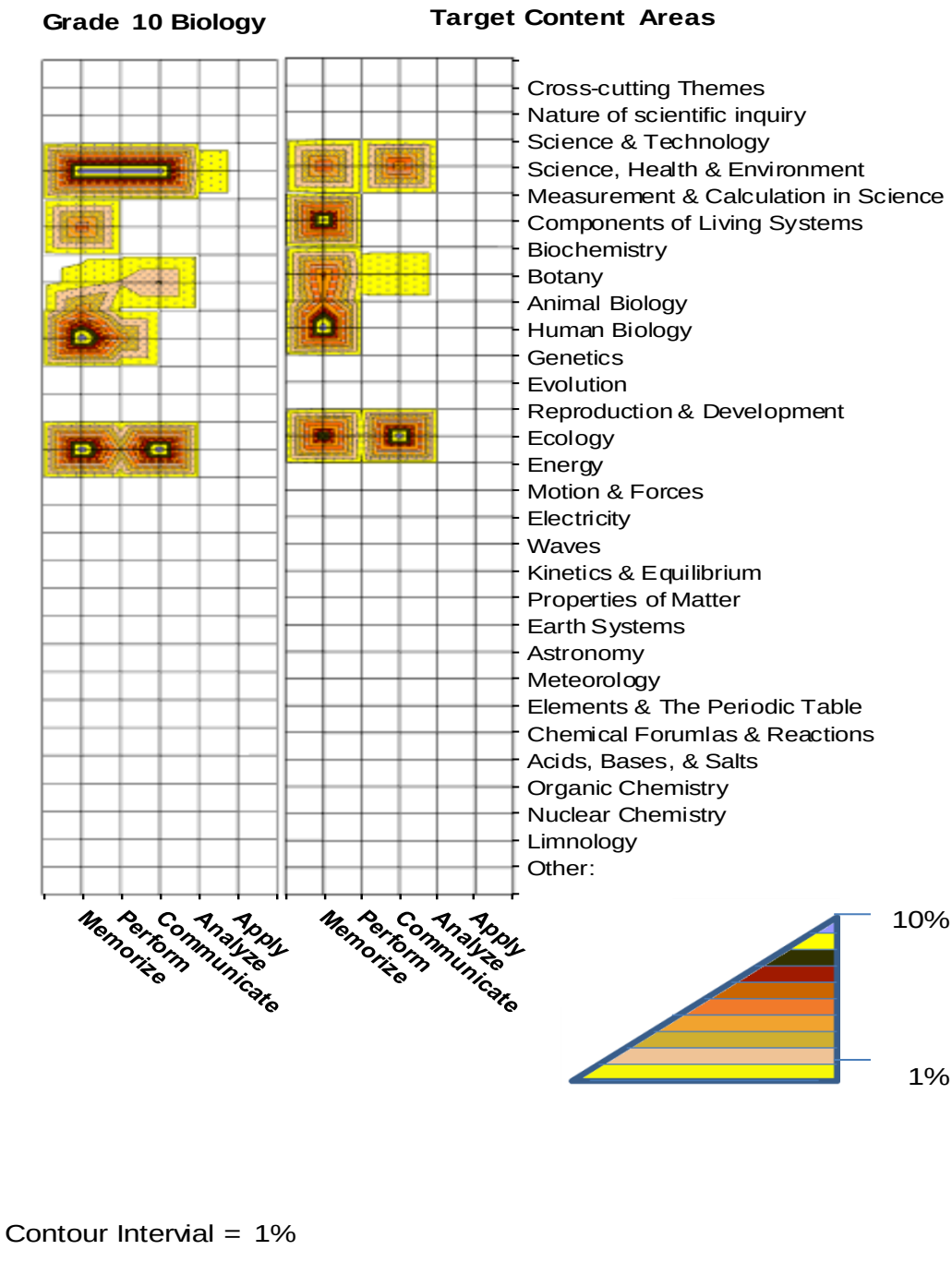
Contour Intervial = 1%

Figure 10.16 Grade 8 **NCEXTEND1** *Science Assessment and Standard Content Map*



Contour Intervial = 1%

Figure 10.17 Grade 10 **NCEXTEND1** *Biology Assessment and Standard Content Map*



10.5 Evidence Regarding Relationships with External Variables

Analysis of the relationship of test scores to variables external to the test provides another important source of convergent and divergent 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 (AERA, APA, & NCME, 2014). Evidence regarding relationships with criteria (i.e., previously termed 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 ***NCEXTEND1*** alternate assessments, teachers' judgments of student achievement levels served as sources of evidence of concurrent validity between the students' actual test score and their predicted performance. The variable Anticipated Achievement Level was provided by the teacher, who worked with the student on the subject during the school year. They were required to review the ***NCEXTEND1*** achievement level descriptors (see Appendix 11-A) before completing their ratings. Teachers then assigned achievement levels that clearly and consistently describes each student base on their professional opinion.

The polychoric correlation coefficient for teacher anticipated achievement level and student's actual achievement level (Table 10.11) ranged from 0.47 to 0.62, a moderate correlation that sufficiently demonstrates that teachers can moderately predict students' achievement level. The complete classification at each achievement level by grade are presented in Appendix 10-A Anticipated and Actual Achievement Level.

Moreover, the correlation coefficients of actual achievement levels between the ELA, math, and science for a given grade were moderate ranging from 0.61 to 0.82 (Table 10.11), suggesting that those who do well in one subject also did reasonably well in other subjects.

Table 10.10 Correlation Coefficient between Teacher Anticipated Achievement Level and Actual Achievement Level in **NCEXTEND1** *2012–2013*

	Anticipated Achieve Level vs. Actual Achieve Level					
	Math		ELA		Science	
	N	Corr	N	Corr	N	Corr
Grade 3	972	0.48	971	0.52		
Grade 4	1,047	0.51	1,045	0.51		
Grade 5	1,076	0.54	1,071	0.58	1,073	0.56
Grade 6	1,129	0.51	1,127	0.54		
Grade 7	1,072	0.50	1,073	0.60		
Grade 8	1,159	0.47	1,157	0.55	1,159	0.55
Grade 10	836	0.58	838	0.62	837	0.56

Table 10.11 **NCEXTEND1** *Correlation Coefficient between Actual Achievement Levels by Content*

Subjects	3	4	5	6	7	8	10
Math and ELA	0.71	0.74	0.72	0.70	0.65	0.64	0.64
Math and Science			0.72			0.61	0.61
ELA and Science			0.82			0.80	0.73

10.6 Fairness and Accessibility

10.6.1 Accessibility in Universal Design

To ensure fairness and accessibility for all eligible students for NC assessments, the principles of universal design were used throughout the development and design of ***NCEXTEND1*** assessments. The ***NCEXTEND1*** assessments measures what students with significant cognitive disability know and are able to do as defined in the North Carolina Extended Content Standards. Assessment must ensure comprehensible access to the content being measured to allow students to accurately demonstrate their standing in the content assessed. In order to ensure items and assessments were developed with universal design principles, NCDPI organized a workshop named "Plain English Strategies: Research, Theory, and Implications for Assessment development" in April 2011. Dr. Edynn Sato who was then Director of Research and English

Learner Assessment at WestEd was invited to train NCDPI test development staff including curriculum staff as well as employees from NC-TOPS on universal design principles and writing in plain English language. The universal design principles were applied in every step of the test development, administration, and reporting.

Evidence of universal design principles applied in the development of ***NCEXTEND1*** assessments (so that students could show what they know) has been documented throughout the item development and review, form review, and test administration sections in the report. Some of the universal design principles applied include:

- Precisely defined constructs
    - Direct match to objective being measured
- Accessible, nonbiased items[12]
    - Accommodations included from the start (Braille, large-print, oral presentation etc.)
    - Ensure that quality is retained in all items
- Simple, clear directions and procedures
    - Presented in understandable language
    - Use simple, high frequency, and compound words
    - Use words that are directly related to content the student is expected to know
    - Omit words with double meanings or colloquialisms
    - Consistency in procedures and format in all content areas
- Maximum legibility
    - Simple fonts
    - Use of white space
    - Headings and graphic arrangement
    - Direct attention to relative importance
    - Direct attention to the order in which content should be considered

[12] See discussions on bias review in Chapter 4

- Maximum readability: plain language
    - Increases validity to the measurement of the construct
    - Increases the accuracy of the inferences made from the resulting data
    - Active instead of passive voice
    - Short sentences
    - Common, everyday words
    - Purposeful graphics to clarify what is being asked

- **Accommodations**
    - One item per page
    - Extended time for ELL Students
    - Test in a separate room

10.6.2 Fairness in Access

As documented throughout Chapter 3, and alignment evidence presented in section 10.4 of this report, the NCDPI ensured that all assessment blueprints are aligned to agree upon content domains which are also aligned to the NC*SCS*. ***NCEXTEND1*** glossary of graphics, Extended Content Standards are published on the NCDPI public website with other relevant information regarding the development of the ***NCEXTEND1*** assessments. This ensures schools and students have exposure to content being targeted in the assessments and thus provides them with an opportunity to learn.

Prior to the administration of the first operational form of the ***NCEXTEND1*** assessments, the NCDPI also published sample release sets with associated administrator scripts and manipulatives for every grade level. These released sets provided students, teachers, and parents with sample items and a sample set of practice items similar to items on the operational assessment. These release sets and scripts also served as a resource to help teachers and students to become familiar with the response formats in the new assessments.

10.6.3 Fairness in Administration

Chapter 5 of this report documents the procedures put in place by the NCDPI to assure the administration of the ***NCEXTEND1*** assessments is uniform, fair, and secured for all students across the state. For each assessment the NCDPI publishes a ***NCEXTEND1*** *Assessment Guide* which is the main training material for all test administrators across the state. These guides provide a comprehensive details of key features about each assessment. Key information provided includes a general overview of each assessment which covers–the purpose of the assessment, eligible students, and testing window and makeup testing options. Assessment guides also covers all preparations and steps that should be followed the day before testing, on test day, and after testing. In addition assessment guides are also used to train test administrators, the NCDPI also publishes a *Proctor's Guide* which is used by test coordinators to train proctors.

One of the purposes for the audits during ***NCEXTEND1*** operational field-test administration window in 2012 was to evaluate the level of adherence to test administration and scoring procedures. Given changes to the assessment's design from previous editions, auditors collected data on student responses, the length of time for administration, and the auditors also provided additional comments regarding the accessibility of test materials for students of varying areas of disability.

To ensure that the implementation of the new design did not result in a negative impact of student's performance, time data were collected and analyzed to make sure there was no impact in student's ability to access the assessment. The results showed an average of 26 minutes for each administration and an overall range of 14 to 59 minutes across the 46 students who participated in this audit

All tests that are part of the North Carolina Testing Program require a standardized process of administration. Review of collected audit data across all schools showed that Assessors followed most test administration directions and scripts, and they paid careful attention to item presentation, maintenance of student engagement throughout the assessment, and evidence of appropriate rapport between Assessors and students. Information regarding any irregularity during administrations was shared with the LEA and school principal during the schools' exit conference calls, and safeguards to avoid future occurrences will be implemented into the ***NCEXTEND1*** test

administrator training process. Results of the audit led to revision of the training process, the *Assessment Guide*, and the Assessor Booklets for ***NCEXTEND1***.

References

AERA, APA, & NCME (2014). *Standards for educational and psychological testing.* Washington, D.C.: Author.

Brennan, R. L. (2004). *Manual for BB-CLASS: A computer program that uses the Beta-Binomial model for classification consistency and accuracy*. Iowa City, IA: Center for Advanced Studies in Measurement and Assessment (CASMA).

Cizek, G. J. (2001). More unintended consequences of high-stakes testing. *Educational Measurement: Issues and Practice*, 20(4), 19-27.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 22*(3), 297-334.

Hanson, B.A. & Brennan, R.L. (1990). An investigation of classification consistency indexes estimated under alternative strong true score models. *Journal of Educational Measurement*, 27(4), 345-359.

Jaeger, R. M. (1991). Selection of judges for standard setting. *Educational Measurement: Issues and Practices*, 10(2), 3-6, 10, 14.

Kane, M. T. (1994). Validating the performance standards associated with passing scores. *Review of Educational Research*, 64, 425–461.

Kane, M. T. (2001). So much remains the same: Conception and status of validation in setting standards. In G. J. Cizek (Ed.), *Setting performance standards: Concepts, methods, and perspectives* (pp. 53–88). Mahwah, NJ: Lawrence Erlbaum.

Linn, R. L. (2002). The measurement of student achievement in international studies. In A. C. Porter & A. Gamoran (Eds). *Methodological Advances in Large-Scale Cross-National Education Surveys* (pp. 25-57). Washington, DC: Board on Testing and Assessment, Center for Education, Division of Behavioral and Social Sciences and Education, National Academy Press.

Livingston, S. A. & Lewis, C. (1995). Estimating the consistency and accuracy of classifications based on test scores. *Journal of Educational Measurement,* 32(2), 179-197.

Plake, B. & Hambleton, R. (2001). The analytic judgment method for setting standards on complex performance assessments. In G. J. Cizek (Ed*.) Setting Performance Standards: Concepts, Methods, and Perspectives* (pp. 283-312). Mahwah, NJ: Lawrence Erlbaum Associates.

Raymond, M. R. & Reid, J. B. (2001). Who made thee a judge? Selecting and training participants for standard setting. In G. J. Cizek (Ed.) *Setting Performance Standards: Concepts, Methods, and Perspectives* (pp. 119-157). Mahwah, NJ: Lawrence Erlbaum & Associates, Inc.

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

Thurlow, M., & Bolt, S. (2001). *Empirical support for accommodations most often allowed in state policy.* (Synthesis Report 41). Minneapolis, MN: National Center on Educational Outcomes. Retrieved [January 25, 2016] from http://www.cehd.umn.edu/nceo/onlinepubs/Synthesis41.html

Webb, N. L. (1997). *Criteria for alignment of expectations and assessments in mathematics and science education* (NISE Research Monograph No. 6). Madison: University of Wisconsin–Madison, National Institute for Science Education. Washington, DC: Council of Chief State School Officers.

Webb, N. L. (2005). *Web Alignment Tool*. Wisconsin Center of Educational Research. University of Wisconsin-Madison. Retrieved [January, 2016] from http://wat.wceruw.org/index.aspx

Appendix 2-A
North Carolina *Testing Code of Ethics*

Appendix 2 - A

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) dissagregation 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 3-A
Norm Webb Training-
Content Complexity

Content Complexity
Norman L. Webb
Wisconsin Center for Education Research
Supported by the National Science Foundation

North Carolina Department of Instruction
Raleigh, North Carolina
July 26, 2010

Outline of Day	Outline of Workshop
Session 1	History of Categorization Schemes for Identifying Content Complexity
Session 2	Depth-of-Knowledge Definitions
Session 3	Depth-of-Knowledge Practicum and the Ins and Outs
Session 4	Alignment of Standards and Assessments

Importance of Content Complexity

- Vastness of Content
- Alignment
- Validity
- Clarity
- Teacher Guidance
- Truth in Advertising

Content Complexity

Differentiates learning expectations and outcomes by considering the amount of prior knowledge, processing of concepts and skills, sophistication, number of parts, and application of content structure required to meet an expectation or to attain an outcome.

Tyler's Behavioral Aspect of the Objectives (course dependent)

1. Understanding of important facts and principles
2. Familiarity with dependable sources of information
3. Ability to interpret data
4. Ability to apply principles
5. Ability to study and report results of study
6. Broad and mature interests
7. Social attitudes

Bloom Taxonomy

Knowledge	Recall of specifics and generalizations; of methods and processes; and of pattern, structure, or setting.
Comprehension	Knows what is being communicated and can use the material or idea without necessarily relating it.
Applications	Use of abstractions in particular and concrete situations.
Analysis	Make clear the relative hierarchy of ideas in a body of material or to make explicit the relations among the ideas or both.
Synthesis	Assemble parts into a whole.
Evaluation	Judgments about the value of material and methods used for particular purposes.

Gagné's Conditions of Learning

- Signal Learning
- Stimulus-Response Learning
- Chaining
- Verbal Association
- Multiple Discrimination
- Concept Learning
- Principle of Learning
- Problem Solving

National Longitudinal Study of Mathematical Abilities (1965-1975)
Model for Mathematics Achievement—Content by Behavior Matrix

	Number Systems	Geometry	Algebra
Computation			
Comprehension			
Application			
Analysis			

NAEP Mathematical Abilities (1990-2005)

Conceptual understanding

Recognize, label, and generate examples of concepts; use & interrelate models, diagrams, manipulatives, & varied representations of concepts; etc.

Procedural knowledge

Select and apply appropriate procedures correctly; verify or justify the correctness of a procedure using concrete models or symbolic methods; or extend or modify procedures to deal with factors inherent in problem settings.

Problem solving

Recognize and formulate problems; determine the consistency of data; use strategies, data, models; generate, extend, & modify procedures; use reasoning in new settings; & judge the reasonableness & correctness of solutions.

U.S. Department of Education Guidelines

Dimensions important for judging the alignment between standards and assessments

- ***Comprehensiveness***: Does assessment reflect full range of standards?
- ***Content and Performance Match***: Does assessment measure what the standards state students should both know & be able to do?
- ***Emphasis***: Does assessment reflect same degree of emphasis on the different content standards as is reflected in the standards?
- ***Depth***: Does assessment reflect the cognitive demand &depth of the standards? Is assessment as cognitively demanding as standards?
- ***Consistency with achievement standards***: Does assessment provide results that reflect the meaning of the different levels of achievement standards?
- ***Clarity for users***: Is the alignment between the standards and assessments clear to all members of the school community?

Survey of Enacted Curriculum Mathematics Cognitive Levels

- Memorize
  Recall basic mathematics facts; etc.
- Perform procedures
  Do computational procedures or algorithms; etc.
- Demonstrate understanding
  Communicate mathematical ideas; use representations to model mathematical ideas; etc.
- Conjecture, generalize, prove
  Determine the truth of a mathematical pattern or proposition; write formal or informal proof; etc.
- Solve non-routine problems, make connections
  Apply and adapt a variety of appropriate strategies to solve problems; etc.

Survey of Enacted Curriculum English Language Arts Cognitive Levels

- Recall
  Provide facts, terms, definitions, conventions; describe; etc.
- Demonstrate/Explain
  Follow instructions; give examples; etc.
- Analyze/investigate
  Categorize, schematize; distinguish fact from opinion; make inferences, draw conclusions; etc.
- Evaluate
  Determine relevance, coherence, logical, internal consistency; test conclusions; etc.
- Generate/create
  Integrate, dramatize; predict probable consequences; etc.

Strands of Mathematical Proficiency (Adding It Up, 2001)

Conceptual understanding
- Comprehension of mathematical concepts, operations, & relations

Procedural fluency
- Skill in carrying out procedures flexibly, accurately, efficiently, & appropriately

Strategic competence
- Ability to formulate, represent, & solve mathematical problems

Adaptive reasoning
- Capacity for logical thought, reflection, explanation, & justification

Productive disposition
- Habitual inclination to see mathematics as sensible, useful, & worthwhile, coupled with a belief in diligence & one's own efficacy (p. 116)

Mathematical Complexity of Items NAEP 2005 Framework

The demand on thinking the items requires:

Low Complexity
- Relies heavily on the recall and recognition of previously learned concepts and principles.

Moderate Complexity
- Involves more flexibility of thinking and choice among alternatives than do those in the low-complexity category.

High Complexity
- Places heavy demands on students, who must engage in more abstract reasoning, planning, analysis, judgment, and creative thought.

Marzano's Dimension of Thinking (Wisconsin DPI) (1989)

- Gathering Information
  - Observe, recall, question
- Organizing Information
  - Represent, compare, classify, order
- Analyzing Information
  - Attributes and components, patterns and relationships, main points, accuracy and adequacy
- Generating Information
  - Infer, predict, elaborate
- Integrating Information
  - Summarize, restructure
- Evaluating Information
  - Establish criteria, verify

Developing Cognitive Complexity Definitions

Depth of Knowledge (1997)

Level 1 Recall
Recall of a fact, information, or procedure.

Level 2 Skill/Concept
Use information or conceptual knowledge, two or more steps, etc.

Level 3 Strategic Thinking
Requires reasoning, developing plan or a sequence of steps, some complexity, more than one possible answer.

Level 4 Extended Thinking
Requires an investigation, time to think and process multiple conditions of the problem.

Which of these means about the same as the word *gauge*?

a. **balance**

b. **measure**

c. **select**

d. **warn**

level 1

A car odometer registered 41,256.9 miles when a highway sign warned of a detour 1,200 feet ahead. What will the odometer read when the car reaches the detour? (5,280 feet = 1 mile)

(a) 42,456.9

(b) 41,279.9

(c) 41,261.3

(d) 41,259.2

(e) 41,257.1

Did you use the calculator on this question?

☐ Yes ☐ No

level 2

$$\begin{array}{r} 121 \\ 13 \\ 32 \\ +\ 34 \\ \hline \end{array}$$

1) 190
2) 200
3) 290
4) N

level 1

Which of these conclusions is best supported by information from the passage?

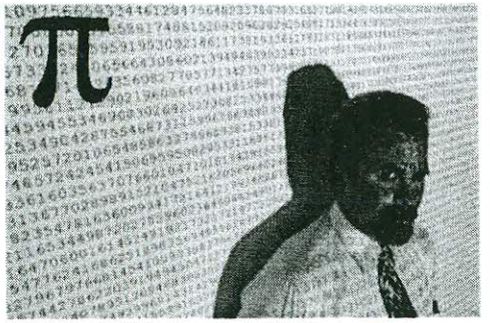
a. If a candidate meets the personal and educational qualifications and is in fair physical shape, his or her chances of becoming an agent are very good.

b. Compared with other law enforcement agencies in the country, the F.B.I. has a low success rate for tracking down and apprehending suspected offenders.

c. The job of an agent is not for everyone; it takes someone with special training who is not afraid of danger and doesn't mind being socially isolated at times.

d. The life of a federal investigator is not as interesting as most people think; agents spend most of their time working at desks.

It is Still A Level 1



Marc Umile poses for a picture in front of a projection of the string of numbers knows as pi in Philadelphia, Friday, March, 2, 2006. Umile is among a group of people fascinated with pi, a number that has been computed to more than a trillion decimal places. He has recited pi to 12,887 digits, perhaps the U.S. record. (AP Photo/Matt Rourke)

Depth of Knowledge Framework for the Wisconsin Knowledge and Concepts Examinations Re-alignment Study

TerraNova Thinking Skill	Descriptor	Depth of Knowledge Levels			
		1—Recall of Information	2—Basic Reasoning	3—Complex Reasoning	4—Extended Reasoning
Gathering Information	Observe	✓			
	Recall	✓			
	Question	✓	✓		
Organizing Information	Represent	✓	✓		
	Compare		✓		
	Classify		✓		
	Order		✓		
Analyzing Information	Attributes & Components	✓	✓		
	Patterns & Relationships		✓		
	Main Points		✓		
	Accuracy & Adequacy		✓		
Generating Information	Infer		✓	✓	
	Predict		✓	✓	
	Elaborate		✓	✓	
Integrating Information	Summarize				
	Restructure		✓	✓	
Evaluating Information	Establish Criteria		✓	✓	
	Verify		✓	✓	

Hess's Bloom's & DOK Levels

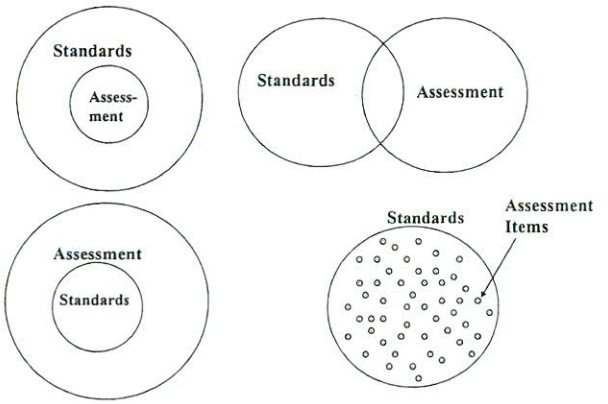
Bloom's Revised Taxonomy of Cognitive Process Dimensions	Webb's Depth-of-Knowledge (DOK) Levels			
	Level 1 Recall & Reproduction	Level 2 Skills & Concepts	Level 3 Strategic Thinking/ Reasoning	Level 4 Extended Thinking
Remember				
Understand				
Apply				
Analyze				
Evaluate				
Create				

Review DOK Definitions and Sample Objectives and Items

Alignment Process

- Identify Standards and Assessments
- Select 6-8 Reviewers (Content Experts)
- Train Reviewers on DOK Levels
- Part I: Code DOK Levels of the Standards/Objectives
- Part II: Code DOK Levels and Corresponding Objectives of Assessment Items

Degree of Alignment



Specific Criteria

Content Focus

A. Categorical Concurrence

B. Depth-of-Knowledge Consistency

C. Range-of-Knowledge Correspondence

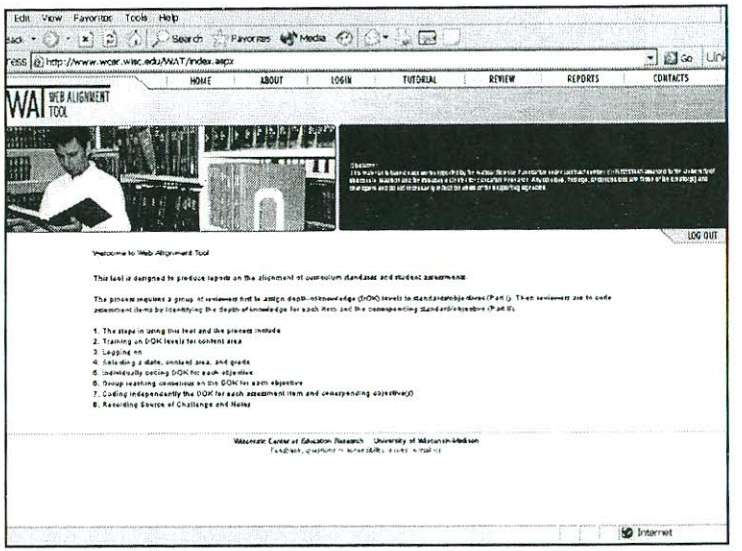
D. Balance of Representation

Alignment Levels Using the Four Criteria

Alignment Level	Categorical Concurrence	Depth of Knowledge	Range of Knowledge	Balance of Representation
Acceptable	6 item per standard	50%	50%	0.70
Weak		40% - 49%	40% - 49%	.60 - .69
Unacceptable	Less than 6 items per standard	Less than 40%	Less than 40%	Less than .60

Coding Process Tips

- One Primary Objective and up to Two Secondary Objectives (if necessary)
- Source of Challenge (a correct/incorrect response for the wrong reason)
- Notes (any insights to share)
- Consider Full Range of Standards
- Use generic objectives sparingly



Subject	Depth of Knowledge			
	Level 1	Level 2	Level 3	Level 4
Mathematics	Requires students to recall or observe facts, definitions, or terms. Involves simple one-step procedures. Involves computing simple algorithms (e.g., sum, quotient).	Requires students to make decisions of how to approach a problem. Requires students to compare, classify, organize, estimate or order data. Typically involves two-step procedures.	Requires reasoning, planning or use of evidence to solve problem or algorithm. May involve activity with more than one possible answer. Requires conjecture or restructuring of problems. Involves drawing conclusions from observations, citing evidence and developing logical arguments for concepts. Uses concepts to solve non-routine problems.	Requires complex reasoning, planning, developing and thinking. Typically requires extended time to complete problem, but time spent not on repetitive tasks. Requires students to make several connections and apply one approach among many to solve the problem. Involves complex restructuring of data, establishing and evaluating criteria to solve problems

Questions for Eliciting Thinking at Different Depth-of-Knowledge Levels

- DOK 1:
  - How can you find the meaning of _____?
  - Can you recall_____?
- DOK 2:
  - How would you classify the type of _____?
  - What can you say about_____?
  - How would you summarize _____?
- DOK 3:
  - What conclusion can be drawn from these three texts _____?
  - What is your interpretation of this text? Support your rationale.

Issues with DOK

Issues in Assigning Depth-of-Knowledge Levels

- Complexity vs. difficulty
- Distribution by DOK Level
- Item type (MS, CR, OE)
- Central performance in objective
- Consensus process in training
- Application to instruction
- Reliabilities

Distribution of Depth-of-Knowledge Levels from Different States
Language Arts

Standard	Number of Objs. Under Standard	DOK Levels of Objs.	# of Objs by DOK Levels	% of Objs by DOK Levels
Michigan High School	55	1 2 3 4	0 15 31 9	0 27 57 16
West Virginia Grade 8	32	1 2 3 4	2 12 16 2	6 37 50 6
Alabama Grade 8	4	1 2 3	1 2 1	25 50 25

Distribution of Depth-of-Knowledge Levels from Different States
Mathematics

	Total Number of Objectives	DOK Level	# of Objs by Level	% within std by Level
Michigan High School	77	1 2 3 4	9 41 24 3	11 53 31 3
West Virginia Grade 8	34.25	1 2 3	4 20 8	12 62 25
Alabama Grade 8	14.75	1 2 3	6 7 1	42 50 7

Common Core Standards

Mathematics

Grade 5 Number and Operations-Fractions

Use equivalent fractions as a strategy to add and subtract fractions.

- 1. Add and subtract fractions with unlike denominators (including mixed numbers) by replacing given fractions with equivalent fractions in such a way as to produce an equivalent sum or difference of fractions with like denominators. *For example, 2/3 + 5/4 = 8/12 + 15/12 = 23/12. (In general, a/b + c/d = (ad + bc)/bd.)*
- 2. Solve word problems involving addition and subtraction of fractions referring to the same whole, including cases of unlike denominators, e.g., by using visual fraction models or equations to represent the problem. Use benchmark fractions and number sense of fractions to estimate mentally and assess the reasonableness of answers. *For example, recognize an incorrect result 2/5 + 1/2 = 3/7 by observing that 3/7 < 1/2.*

Grade 5 Number and Operations--Fractions

4. Apply and extend previous understandings of multiplication to multiply a fraction or whole number by a fraction.

   a. Interpret the product $(a/b) \times q$ as a parts of a partition of q into b equal parts; equivalently, as the result of a sequence of operations $a \times q \div b$. *For example, use a visual fraction model to show* $(2/3) \times 4 = 8/3$, *and create a story context for this equation; do the same with* $(2/3) \times (4/5) = 8/15$. (*In general,* $(a/b) \times (c/d) = ac/bd$.)

   b. Find the area of a rectangle with fractional side lengths by tiling it, and show that the area is the same as would be found by multiplying the side lengths; multiply fractional side lengths to find areas of rectangles, and represent fraction products as rectangular areas.

Reading Standards for Literature K–5
Grade 5

1. Quote accurately from a text when explaining what the text says explicitly and when drawing inferences from the text.
2. Determine a theme of a story, drama, or poem from details in the text, including how characters in a story or drama respond to challenges or how the speaker in a poem reflects upon a topic; summarize the text.
3. Compare and contrast two or more characters, settings, or events in a story or drama, drawing on specific details in the text (e.g., how characters interact).

Web Sites

http://facstaff.wcer.wisc.edu/normw/

Alignment Tool

http://www.wcer.wisc.edu/WAT

Appendix 3-B
Exhibit 307 Plain English Training_042811

Appendix 3-B Exhibit 307 Plain English Training_042811

Hope Lung

Subject: Plain English Strategies Workshop
Location: Room 150

Start: Thu 4/28/2011 8:30 AM
End: Thu 4/28/2011 4:00 PM

Recurrence: (none)

Meeting Status: Meeting organizer

Organizer: Audrey Martin-McCoy

As previously announced, the plain English strategies workshop will be held on April 28. Attached you will find a draft agenda for the day.

The workshop will be held in room 150 of the Education Building, 8:30 am - 4:00 pm.

Audrey

Audrey Martin-McCoy, Ph.D.
Education Testing/Accountability Consultant
Testing Policy and Operations Section/Division of Accountability Services
North Carolina Department of Public Instruction
6314 Mail Service Center
Raleigh, NC 27699-6314

>>> Audrey Martin-McCoy 03/16/11 11:22 AM >>>
A workshop will be offered in an attempt to extend and refine our knowledge and use of plain English language practices in test construction. The workshop will be facilitated by Dr. Edynn Sato. Edynn is Director of Research and English Learner Assessment with the Assessment and Standard Development Services Program at West Ed. She is also the Director of Special Populations at the Assessment and Accountability Comprehensive Center at West Ed.

The training workshop will focus on the latest research in the area of plain English practices and examine its use in our current training used for our item writers/editors and in released state test forms. In sum, this is an opportunity to build and/or re-evaluate how we go about developing plain English test items. Follow up conference calls will be scheduled after the workshop to foster continued understanding of concepts discussed.

The workshop will be held on April 28, 2011, from 8:30 am to 4:00 pm in room 150 at the Education Building. Lunch is on your own from 11:30 am to 12:30 pm. A draft agenda will be sent within the next two weeks. Personnel from DPI ESL, Accountability, and NCSU - TOPS will be invited to attend.

Please save this date and time. Let me know if you have questions.

Audrey

Assessment and Accountability Comprehensive Center
AACC

Appalachia Regional Comprehensive Center at EDVANTIA

WORKSHOP

Plain English Strategies:
Research, Theory, and Implications for Assessment Development

Agenda

April 28, 2011

Workshop Objective: To provide participants with information about plain English strategies that will inform and support the effective application of these practices in the state's test item development process.

8:30 — 8:45 am	Welcome and Introductions *Shirley Carraway, ARCC- NC Liaison* *Audrey Martin-McCoy, NCDPI*
8:45 — 10:00 am	Introduction to Plain English: Research, Theory, and the Accessibility Context *Edynn Sato, AACC- WestEd* Director *Rachel Lagunoff, AACC – WestEd*
10:00 — 10:15 am	Break
10:15 — 11:30 am	Introduction to Plain English: Research, Theory, and the Accessibility Context (Continued) *Edynn Sato and Rachel Lagunoff*
11:30 am — 12:30 pm	Lunch
12:30 pm — 3:30 pm	Application of Plain English Strategies: Implications for Item Development and Related Training *Edynn Sato and Rachel Lagunoff*
3:30 pm — 4:00 pm	Discussion of Possible Next Steps *NCDPI Staff*

Plain English Strategies
Application of Plain English Strategies: Implications for Item Development

WORKSHOP

Examples of applying research-based Plain English strategies to test items

Research Findings	Practical Recommendations	Examples
Words that are short (simple morphologically) tend to be more familiar and, therefore, easier.	Use simple words; use high-frequency words; only use compound words and words with prefixes or suffixes that are likely to be familiar. Exception: words that are directly related to content the student is expected to know	Change *utilize* to *use* Even though *chair* is EDL 2 and *man* is EDL 1, *chairman* is EDL 7, so may not be familiar; both *base* and *baseball* are EDL 3, so likely to be equally familiar. *Proper* is EDL 5, but *improper* is EDL 8, so *im-* is likely to be an unfamiliar prefix; *happy* is EDL 1, and *unhappy* is EDL 2, so *un-* is likely to be a familiar prefix.
Passages with words that are familiar (simple semantically) are easier to understand.	Use familiar words. Omit or define words with double meanings or colloquialisms.	Change *go off* to *leave*, *explode*, or *start to ring* Even seemingly simple words can have multiple meanings, e.g., *fine* (feeling, weather, hair or line, penalty, etc.). Even seemingly simple words can have colloquial or idiomatic uses, e.g., *hop in, blow up, get it.*

Research Findings	Practical Recommendations	Examples
Longer sentences tend to be more complex syntactically and, therefore, more difficult to comprehend.	Retain Subject-Verb-Object structure for statements. Begin questions with question words. Avoid clauses and phrases.	Change *At which of the following times* to *When* Change *A report that contains 64 papers* to *He needs 64 sheets of paper for each report*
Long items tend to pose greater difficulty.	Remove unnecessary expository material.	Change *The weights of four different bookbags are recorded in the chart above. According to the chart, which bookbag is the heaviest?* to *Look at the chart below. Which bookbag weighs the MOST?*
Complex sentences tend to be more difficult than simple or compound sentences.	Keep to the present tense, use active voice, avoid the conditional mode, and avoid starting with sentence clauses.	Change *The weights of 3 objects were compared* to *Sandra compared the weights of 3 objects* Change *If Lee delivers x newspapers* to *Lee delivers x newspapers*

Suggested Strategies for Ensuring Maximum Test Item Readability and Comprehensibility

Strategy	Example
Avoid irregularly spelled words	Words such as *trough* or *feign* may be difficult to read
Use generic terms and familiar proper names with simple spelling	Use *tree* instead of *pine* or *oak*; use *Jeff* instead of *Geoffrey* and *Ellen* instead of *Eleanor*
Avoid multiple terms for the same concept	Do not use both *children* and *kids* in an item or a set of items; in items based on a reading passage, use the same term as in the passage
Make sure all noun-pronoun relationships are clear	In the stem *Scientists think bears are most dangerous when they are*, replace *they* with *the bears*
Put important context first	When time and setting are important to the sentence, place them at the beginning of the sentence; put the location of information in a passage at the beginning of the stem (e.g., *In the 1800s*; *In the second paragraph*)
When possible, write closed stems that end with a question mark	If the answer choices are complete sentences, a closed stem is usually possible; if words are repeated at the beginning of answer choices, an open stem may be preferable

References

Abedi, J. et al. (2005). *Language Accommodations for English Learners in Large-Scale Assessments: Bilingual Dictionaries and Linguistic Modification.* (CSE Report 666). Los Angeles: University of California, Center for the Study of Evaluation/National Center for Research on Evaluation, Standards, and Student Testing.

Brown, P.J. (1999). *Findings of the 1999 plain language field test*. University of Delaware, Newark, DE: Delaware Education Research and Development Center.

Gaster, L., & Clark, C. (1995). *A guide to providing alternate formats.* West Columbia, SC: Center for Rehabilitation Technology Services. (ERIC Document No. ED 405689)

Thompson, S. J., Johnstone, C. J., & Thurlow, M. L. (2002). *Universal design applied to large scale assessments* (Synthesis Report 44). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Retrieved April 25, 2011, from the World Wide Web: http://education.umn.edu/NCEO/OnlinePubs/Synthesis44.html

Evaluating Items for Plain English: Sample Items

SAMPLE A
Reading Comprehension
Grade 3
Selection: *Hamish McBean and His Sheep*

2. Which words from the selection ***best*** help the reader picture the setting?

SAMPLE B
Reading Comprehension
Grade 3
Selection: *Lots of Kids Live Here*

9. Which completes the chart?

kids	young goats
does	female goats
bucks	?

A old goats

B male goats

C mother goats

D newborn goats

SAMPLE C
Reading Comprehension
Grade 5
Selection: *Seneca Oil and Early America*

18. According to the selection, what was one effect of the Senecas' mixing petroleum with paint, particularly during a time of war?

SAMPLE D
Reading Comprehension
Grade 8
Selection: *Here's to Ears*

15. Why is impaired hearing called "auditory isolation"?

A It has a single cause.

B It does not involve other body systems.

C It cuts people off from their environment.

D It keeps sound waves from reaching the auditory nerve.

SAMPLE E
Mathematics—Calculator Inactive
Grade 3

2. There are 20 seeds in a package. If 5 seeds are put in each flower pot, how many flower pots are needed to plant all of the seeds?

SAMPLE F
Mathematics—Calculator Active
Grade 4

17. The bread truck makes deliveries to a store 3 days each week. Each delivery has 45 loaves of bread. Which expression could be used to determine the number of loaves of bread delivered in 5 weeks?

SAMPLE G
Mathematics—Calculator Active
Grade 6

29. Marsha wants to find out how other students at her school get to school each day. Which of the following groups, if surveyed, would give her the ***most accurate*** sample of the student body?

SAMPLE H
Algebra I

44. A computer is purchased for $1,200 and depreciates at $140 per year. Which linear equation represents the value, V, of the computer at the end of t years?

Language for Achievement—A Framework for Academic English Language

Handout description:
The *Language for Achievement Framework* (page 2) is theory and research based, and aspects of the framework have been used in the evaluation and development of English language proficiency (ELP) standards and assessments in a number of states, as well as in examinations of linkage or correspondence between state ELP and academic content standards (i.e., to identify aspects of English language needed to facilitate student access to and meaningful engagement with academic content).

This handout also includes a *taxonomy* (page 3) that focuses on academic language functions (as opposed to, for example, social language and linguistic skills) that is intended to serve for the language domain the role that Bloom's taxonomy, for example, serves for the cognitive domain—Bloom's taxonomy serves as a classification system for thinking behaviors that are important to the learning process (Forehand, 2005; Hancock, 1994; Kreitzer & Madaus, 1994; Seddon, 1978). The taxonomy provides a structure for arranging content learning objectives according to the academic language necessary for students to meet a content objective, or set of related objectives. The taxonomy can inform the development of *language progressions* which place the academic language skills and knowledge of the taxonomy on a developmental continuum, reflecting a progression from the most basic and foundational English language skills and knowledge to the most advanced and developed language skills and knowledge relevant to accessing and achieving rigorous academic content. Therefore, the taxonomy has important implications for instructional practices that can support the language related to academic achievement not only of EL students but of *all* students working to meet more rigorous and higher academic expectations.

Also associated with the framework are rubrics related to language complexity (pages 4-6). The language demands represented in the framework (i.e., academic vocabulary and grammar, functions, spoken and written text, classroom discourse) interact with language complexity.

Information presented in this handout is intended for the following purposes:

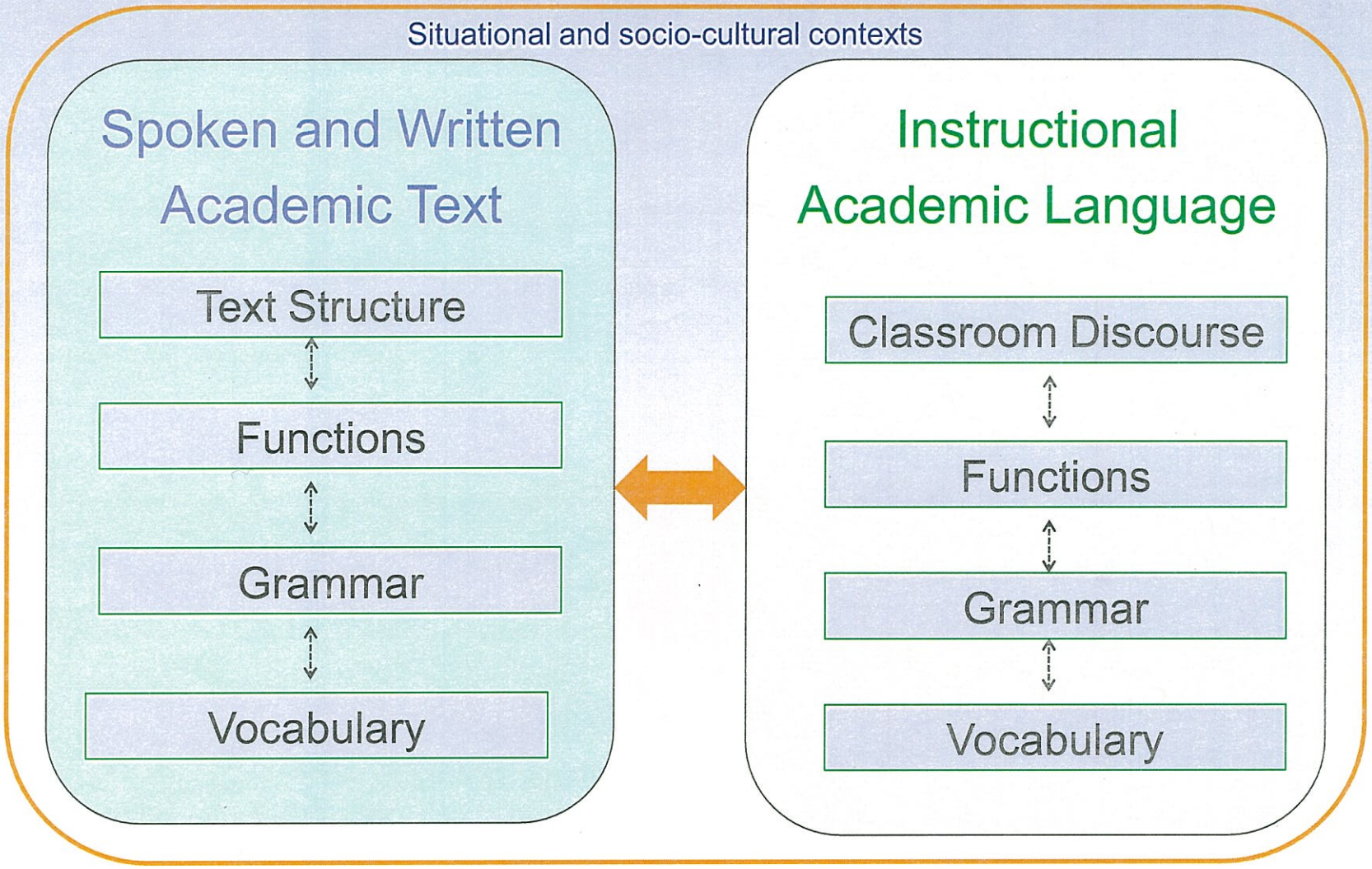
- to help analyze the content and language in standards, assessment tasks, and instructional materials;
- to help make explicit the expectations (cognitive, language) of students;
- to help inform instructional planning and practice so that they are intentional and appropriate in supporting students' progress (cognitive, linguistic) toward proficiency and achievement; and
- to serve as a tool for cross-disciplinary discussions related to appropriately addressing the content and language needs of English learner students and facilitating their achievement in school.

For more information, please contact Dr. Edynn Sato at WestEd (esato@wested.org; 415-615-3226).

Notes:

- For use and distribution of information contained in this packet, please contact Dr. Edynn Sato (contact information listed above).
- The information in this handout was originally developed for research purposes. The information is not necessarily comprehensive (e.g., list of functions).

Language for Achievement: Overview



Additional considerations include: receptive (listening, reading) and productive (speaking, writing) language; language complexity

Contact: Dr. Edynn Sato at esato@wested.org

Language for Achievement—Taxonomy: Academic English Language Functions

Academic English Language Function		*Operational Definition—The language needed to **engage with and achieve** in the content (standard or item) consists of the use of:*
A	Identification	a word or phrase to name an object, action, event, idea, fact, problem, need, or process.
	Labeling	a word or phrase to name an object, action, event, or idea.
	Enumeration	words or phrases to name distinct objects, actions, events, or ideas in a series, set, or in steps.
B	Classification	words, phrases, or sentences to assign/associate an object, action, event, or idea to the category or type to which it belongs.
	Sequencing	words, phrases, or sentences to express the order of information (e.g., a series of objects, actions, events, ideas). Discourse markers include adverbials such as *first, next, then, finally*.
	Organization	words, phrases, or sentences to express relationships between/among objects, actions, events, or ideas, or the structure or arrangement of information. Discourse markers include coordinating conjunctions such as *and, but, yet, or*, and adverbials such as *first, next, then, finally*.
C	Comparison/ Contrast	words, phrases, or sentences to express similarities and/or differences, or to distinguish between two or more objects, actions, events, or ideas. Discourse markers include coordinating conjunctions *and, but, yet, or*, and adverbials such as *similarly, likewise, in contrast, instead, despite this*.
D	Inquiring	words, phrases, or sentences to solicit information (e.g., *yes-no* questions, *wh*-questions, statements used as questions).
E	Description	word, phrase, or sentence to express or observe the attributes or properties of an object, action, event, idea, or solution.
F	Definition	word, phrase, or sentence to express the meaning of a given word, phrase, or expression.
G	Explanation	phrases or sentences to express the rationale, reasons, causes, or relationships related to one or more actions, events, ideas, or processes. Discourse markers include coordinating conjunctions *so, for*, and adverbials such as *therefore, as a result, for that reason*.
H	Retelling	phrases or sentences to relate or repeat information. Discourse markers include coordinating conjunctions such as *and, but*, and adverbials such as *first, next, then, finally*.
	Summarization	phrases or sentences to express important facts or ideas and relevant details about one or more objects, actions, events, ideas, or processes. Discourse structures include: beginning with an introductory sentence that specifies purpose or topic.
I	Interpretation	phrases, sentences, or symbols to express understanding of the intended or alternate meaning of information.
J	Analyzing	phrases or sentences to indicate parts of a whole and/or the relationship between/among parts of an action, event, idea, or process. Relationship verbs such as *contain, entail, consist of*, partitives such as *a part of, a segment of*, and quantifiers such as *some, a good number of, almost all, a few, hardly any* often are used.

Academic English Language Function		*Operational Definition—The language needed to **engage with and achieve** in the content (standard or item) consists of the use of:*
K	Generalization	phrases or sentences to express an opinion, principle, trend, or conclusion that is based on facts, statistics, or other information, and/or to extend that opinion/principle/etc. to other relevant situations/contexts/etc.
	Inferring	words, phrases, or sentences to express understanding of implied/implicit based on available information. Discourse markers include inferential logical connectors such as *although, while, thus, therefore*.
	Prediction	words, phrases, or sentences to express an idea or notion about a future action or event based on available information. Discourse markers include adverbials such as *maybe, perhaps, obviously, evidently*.
	Hypothesizing	phrases or sentences to express an idea/expectation or possible outcome based on available information. Discourse markers include adverbials such as *generally, typically, obviously, evidently*.
L	Argumentation	phrases or sentences to present a point of view with the intent of communicating or supporting a particular position or conviction. Discourse structures include expressions such as *in my opinion, it seems to me*, and adverbials such as *since, because, although, however*.
	Persuasion	phrases or sentences to present ideas, opinions, and/or principles with the intent of creating agreement around or convincing others of a position or conviction. Discourse markers include expressions such as *in my opinion, it seems to me*, and adverbials such as *since, because, although, however*.
	Negotiation	phrases or sentences to engage in a discussion with the purpose of creating mutual agreement from two or more different points of view.
M	Synthesizing	phrases or sentences to express, describe, or explain relationships among two or more ideas. Relationship verbs such as *contain, entail, consist of*, partitives such as *a part of, a segment of*, and quantifiers such as *some, a good number of, almost all, a few, hardly any* often are used.
N	Critiquing	phrases or sentences to express a focused review or analysis of an object, action, event, idea, or text.
O	Evaluation	phrases or sentences to express a judgment about the meaning, importance, or significance of an action, event, idea, or text.
P	Symbolization & Representation	symbols, numerals, and letters, to represent meaning within a conventional context (e.g., +, -, CO_2, >, Δ, π, cos, y=3x+4, $c^2=a^2+b^2$, $h/2(b_1+b_2)$, *cat* vs. cat).
Z	No Academic Language Function	Item or standard does not contain *any* academic language functions; may contain linguistic skills (e.g., phonemic awareness, syllabication).

Note: This taxonomy focuses on academic language functions and does not address the identification or definition of linguistic skills (e.g., phonology, morphology).

Language for Achievement—Language Complexity

The *Language for Achievement* language demands (i.e., academic vocabulary and grammar, functions, spoken and written text, classroom discourse) interact with language complexity. Language complexity, as used in this framework, is defined below.

Vocabulary and Grammar

Lower Complexity	Higher Complexity
• Semantically simple words and phrases • Common, high-frequency words and phrases • Simple, high-frequency morphological structures (e.g., common affixes, common compound words)	• Semantically complex words and phrases (e.g., multiple-meaning words, idioms, figurative language) • Specialized or technical words and phrases • Complex, higher level morphological structures (e.g., higher level affixes and compound words)
• Short, simple sentences with limited modifying words or phrases • SVO sentence structure; simple verb and noun phrase constructions • Simple, familiar modals (e.g., *can*) • Simple *wh-* and *yes/no* questions • Direct (quoted) speech • Verbs in present tense, simple past tense, and future with *going to* and *will* • Simple, high-frequency noun, adjective, and adverb constructions	• Compound and complex sentences; longer sentences with modifying words, phrases, and clauses • High level phrase and clause constructions (e.g., passive constructions, gerunds and infinitives as subjects and objects, conditional constructions) • Multiple-meaning modals, past forms of modals • Complex *wh-* and *yes/no* question constructions, tag questions • Indirect (reported) speech • Present, past, and future progressive and perfect verb structures • Complex, higher level noun, adjective, and adverb constructions

Functions

Lower Complexity	Higher Complexity
• Length ranges from a word to paragraphs • No/little variation in words and/or phrases in sentences/paragraphs; consistent use of language • Repetition of key words/phrases/sentences *reinforces* information • Language is used to present critical/central details • No/little abstraction; language reflects more literal/concrete information; illustrative language is used; language is used to define/explain abstract information • Graphics and/or relevant text features reinforce critical information/details • Mostly common/familiar words/phrases; no/few uncommon words/phrases, compound words, gerunds, figurative language, and/or idioms • Language is organized/structured • Mostly simple sentence construction • No/little passive voice • Little variation in tense • Mostly one idea/detail per sentence • Mostly familiar construction (e.g., 's for possessive; s and es for plural) • Mostly familiar text features (e.g., bulleted lists, bold face)	• Length ranges from a word to paragraphs • Some variation in words and/or phrases in sentences/paragraphs • Repetition of key words/phrases/sentences *introduces new* or *extends* information • Language is used to present critical/central details, but non-essential detail also is presented • Some abstraction; language *may or may not* be used to define/explain abstract information; illustrative language *may or may not* be used; technical words/phrases are used • Graphics and/or relevant text features *may or may not* reinforce critical information/details • Some common/familiar words/phrases; some uncommon words/phrases, compound words, gerunds, figurative language, and/or idioms • Language *may or may not* be organized/structured • Varied sentence construction, including complex sentence construction • Some passive voice • Variation in tense • Multiple ideas/details per sentence • Some less familiar/irregular construction • Some less familiar text features (e.g., pronunciation keys, text boxes)

Spoken and Written Texts

Lower Complexity	Higher Complexity
• Short texts, or longer texts chunked into short sections (words, phrases, single sentences, short paragraphs) • No or little variation of words/phrases in sentences/paragraphs • Repetition of key words/phrases reinforces information • One idea/detail per sentence; only critical/central ideas included • No or little abstraction; mostly literal/concrete information; abstract information is defined or explained • Visual aids, graphics, and/or text features reinforce critical information/details • Common text features (e.g. bulleted lists, boldface font)	• Long texts (long lists of words/phrases, a series of sentences, long paragraphs, multiple-paragraph texts) • Variation of words/phrases in sentences/paragraphs • Repetition of key words/phrases introduces new information or extends information • Multiple ideas/details per sentence; non-essential ideas included • Some or much abstraction that is not explicitly defined or explained • Visual aids, graphics, and/or text features may not reinforce critical information/details • Higher level text features (e.g., pronunciation keys, text boxes)

Classroom Discourse

Lower Complexity	Higher Complexity
• Semantically simple words and phrases • Common, high-frequency words and phrases • Simple, high-frequency morphological structures (e.g., common affixes, common compound words)	• Semantically complex words and phrases (e.g., multiple-meaning words, idioms, figurative language) • Specialized or technical words and phrases • Complex, higher level morphological structures (e.g., higher level affixes and compound words)
• Short, simple sentences with limited modifying words or phrases • SVO sentence structure; simple verb and noun phrase constructions • Simple, familiar modals (e.g., can) • Simple wh- and yes/no questions • Direct (quoted) speech • Verbs in present tense, simple past tense, and future with going to and will • Simple, high-frequency noun, adjective, and adverb constructions Note: To the extent that spoken "texts" (planned, connected utterances) are used in classroom discourse, elements of lower complexity spoken text, as defined previously, apply here.	• Compound and complex sentences; longer sentences with modifying words, phrases, and clauses • High level phrase and clause constructions (e.g., passive constructions, gerunds and infinitives as subjects and objects, conditional constructions) • Multiple-meaning modals, past forms of modals • Complex wh- and yes/no question constructions, tag questions • Indirect (reported) speech • Present, past, and future progressive and perfect verb structures • Complex, higher level noun, adjective, and adverb constructions Note: To the extent that spoken "texts" (planned, connected utterances) are used in classroom discourse, elements of higher complexity spoken text, as defined previously, apply here.

Definition from the *Framework for High-Quality ELP Standards and Assessments* (AACC, 2009):
Academic language, broadly defined, includes the language students need to meaningfully engage with academic content within the academic context. This should *not* be interpreted to suggest that separate word lists and/or definitions of content-related language should be developed for each academic subject. Rather, academic language includes the words, grammatical structures, and discourse markers needed in, for example, describing, sequencing, summarizing, and evaluating — these are language demands (skills, knowledge) that facilitate student access to and engagement with grade-level academic content. These academic language demands are different from cognitive demands (e.g., per Bloom's taxonomy). Although there may not be just one accepted definition of academic language, there are a good number of resources available that address the issue of academic language and may be considered in the development of state ELP standards and assessments. For example: Aguirre-Munoz, Parks, Benner, Amabisca, & Boscardin, 2006; Bailey, 2007; Bailey, Butler, & Sato, 2007; Butler, Bailey, Stevens, Huang, & Lord, 2004; Chamot & O'Malley, 1994; Cummins, 1980; Cummins, 2005; Halliday, 1994; Sato, 2007; Scarcella & Zimmerman, 1998; Schleppegrell, 2001.

For a free download of the *Framework for High-Quality ELP Standards and Assessments*, go to http://www.aacompcenter.org/cs/aacc/print/htdocs/aacc/resources_sp.htm.

From: http://ies.ed.gov/ncee/edlabs/projects/project.asp?ProjectID=92

Accommodations for English Language Learner Students:
The Effect of Linguistic Modification of Math Test Item Sets

Edynn Sato, Stanley Rabinowitz, Carole Gallagher, and Chun-Wei Huang

REL West's study on middle school math assessment accommodations found that simplifying the language—or linguistic modification—on standardized math test items made it easier for English Language learners to focus on and grasp math concepts, and thus was a more accurate assessment of their math skills.

The results contribute to the body of knowledge informing assessment practices and accommodations appropriate for English language learner students.

The study examined students' performance on two sets of math items—both the originally worded items and those that had been modified. Researchers analyzed results from three subgroups of students—English learners (EL), non-English language arts proficient (NEP), and English language arts proficient (EP) students.

Key results include:

- Linguistically modifying the language of mathematics test items did not change the math knowledge being assessed.
- The effect of linguistic modification on students' math performance varied between the three student subgroups. The results also varied depending on how scores were calculated for each student.
- For each of the four scoring approaches analyzed, the effect of linguistic modification was greatest for EL students, followed by NEP and EP students.

Note: The following pages are excerpted from the full report which is available at: http://ies.ed.gov/ncee/edlabs/projects/project.asp?ProjectID=92

ies NATIONAL CENTER FOR EDUCATION EVALUATION AND REGIONAL ASSISTANCE
Institute of Education Sciences

REL WEST

NCEE 2009–4079

U.S. DEPARTMENT OF EDUCATION

Accommodations for English Language Learner Students: the Effect of Linguistic Modification of Math Test Item Sets

Final Report

ies NATIONAL CENTER FOR EDUCATION EVALUATION AND REGIONAL ASSISTANCE
Institute of Education Sciences

REL WEST
Regional Educational Laboratory
At WestEd

REL WEST
Regional Educational Laboratory At WestEd

Accommodations for English Language Learner Students: The Effect of Linguistic Modification of Math Test Item Sets

June 2010

Authors:

Edynn Sato, Principal Investigator
WestEd

Stanley Rabinowitz, Principal Investigator
WestEd

Carole Gallagher, Senior Research Associate
WestEd

Chun-Wei Huang, Senior Research Analyst
WestEd

Project Officer:

Ok-Choon Park
Institute of Education Sciences

NCEE 2009-4079
U.S. Department of Education

ies NATIONAL CENTER FOR EDUCATION EVALUATION AND REGIONAL ASSISTANCE
Institute of Education Sciences

Contents

Tables

Figures

Appendix D. Guide for developing a linguistically modified assessment

[This guide was followed to linguistically modify the items used in this study. Experts in mathematics, linguistics, measurement, curriculum and instruction, and the English language learner student population were convened to discuss linguistic modification strategies and their application. These experts possessed advanced degrees (such as an M.A. or Ph.D.), had classroom teaching experience, and assessment development experience. The selection of items, the linguistic modification of items, and the creation of the item sets used in this study occurred over the equivalent of a period of approximately three weeks and followed generally accepted item development procedures including verification of content alignment, appropriateness for the student population, and freedom from bias and sensitivity issues.]

For all students, access to test content is necessary to ensure the validity of assessment results.[35] Valid assessments are especially critical if results are used to inform classroom instruction or for accountability purposes. When access is constrained in some way (for example, linguistically or cognitively), students may be prevented from fully demonstrating what they know and can do, and the test score may underestimate or misrepresent students' achievement. To assess English language learner students' knowledge of academic content, it is critical to determine whether their academic performance reflects their understanding of the targeted content or their lack of English language proficiency. There is an interaction between how assessed content is presented in test items and what English language learner students need in order to access that content. This interaction affects the validity of the assessment results and the interpretation of those results.

Linguistic modification of test items is an approach for addressing the particular access needs of English language learner students so that test performance is attributable less to English language proficiency and more to knowledge and skills related to the tested content. The approach outlined below is intended to help researchers in this study consider key characteristics of the content and the student population as they develop linguistically modified test items. The three steps in this process are:

- Define the domain and constructs of tested content.
- Define the English language learner population that will be tested.
- Apply and evaluate linguistic modification strategies to test items.

[35] Information in this appendix is drawn from Sato (2008).

Step 1: define the domain and constructs

Articulate the purpose of the assessment. Consider the range of ways the assessment results will be used and the intended outcomes of testing.

Recommended specialists for this step

Given the purpose of the assessment and the population assessed, this step is best conducted by a team that includes content specialists, assessment specialists, curriculum and instruction specialists, English language development specialists, and population specialists (that is, individuals with specialized knowledge about the English language learner student population).

Purpose

The assessment results will be used for the following purpose(s):

_____

Assessed academic content domain

The assessment will measure students' knowledge of:

_____

_____

Considerations
Is this test appropriate for the target content domain? To what degree do content domain characteristics align with the intended purpose of this assessment?

Assessed constructs—content and skills

More specifically, the assessment will measure the following constructs (content and skills) related to the domain:

_____

_____

Considerations
Do the content and skills assessed in the set of linguistically modified test items reflect the intended breadth, depth, and range of complexity of the assessed domain? Are the verbs used in the state standards statements specific enough to guide assessment development (for example, "identify," "describe," "compare" vs. the more vague "know," "understand")? If the latter, how are students expected to demonstrate their knowledge and skills?

Content-related language—language demands

The following language demands are associated with the content and skills that will be assessed (see tables E1 and E2 in appendix E for a list of language demands—linguistic skills and academic language functions):

_____

_____

Considerations
Have students' linguistic skills and academic language functions both been considered?
Is the range of language demands in the linguistically modified items consistent with the breadth, depth, and range of complexity of the assessed content domain?

Content-related language—specific vocabulary and terminology

The following vocabulary and terminology are specific to the grade-level content assessed; therefore, they should not be linguistically modified:

_____

_____

Considerations
Is the vocabulary and terminology identified consistent with the intent of the grade-level content standards?

Step 2: define the population and student subgroups

Articulate the key characteristics and access needs of the English language learner student population. Since this group of students is especially diverse and heterogeneous, it may be necessary to identify key subgroups of students within the state.

Recommended specialists for this step

Given the purpose of the assessment and the population assessed, this step is best conducted by a team that includes content specialists, assessment specialists, curriculum and instruction specialists, English language development specialists, and population specialists (that is, individuals with specialized knowledge about English language learner students).

Student population

The target English language learner population can be characterized as follows
(see appendix E for a description of English language learner students):

Student access needs

Document the access needs of the target English language learner student population, taking into account characteristics such as:

Context
What topics, themes, locations, situations, illustrations, and such are familiar to these students?

Words, phrases, sentences
What written vocabulary is familiar to these students? What phrases are familiar to these students? What sentence structures are familiar to these students? What tenses (for example, present, past) and constructions (for example, plural _s, possessive _'s) are familiar to these students? What proper nouns are familiar to students as a result of their classroom reading?

Format/Style
With what formats/styles are these students familiar (for example, bulleted lists, text boxes, underlining for emphasis)? How is information typically presented to these students during instruction?

Step 3: apply and evaluate linguistic modification strategies

Determine which content and item types lend themselves to linguistic modification. Then develop and evaluate each test item according to the following dimensions: context, graphics, vocabulary/wording, sentence structure, and format/style (see table D1 for linguistic modification guidelines and strategies for each dimension).

Recommended specialists for this step

This step is best conducted by a team that includes content specialists, assessment specialists, curriculum and instruction specialists, English language development specialists, and population specialists (that is, individuals with specialized knowledge of the English language learner population).

Categorize target content and item types

Sort content/test items into one of the following three categories of eligibility for linguistic modification. Within each eligibility category, group content standards and test items by content strand (for example, measurement or algebra for mathematics).

- Definitely eligible.
- Definitely not eligible.
- Possibly eligible.

Considerations
A test item's appropriateness for linguistic modification is associated with the quantity of construct-irrelevant language in that test item; the greater the quantity of construct-irrelevant language, the greater the likelihood that the item can be linguistically modified effectively for English language learner students. There also is a greater likelihood that construct-irrelevant language can be linguistically modified without significantly changing the assessed construct (for example, mathematics achievement).

Apply linguistic modification guidelines and strategies

For content/items that are eligible and possibly eligible for linguistic modification, systematically apply the relevant guidelines and strategies presented in table D1 (that is, context, graphics, vocabulary/wording, sentence structure, format/style).

Considerations
The team of specialists who are linguistically modifying items need specialized training to ensure that they are appropriately applying linguistic modification guidelines. It is important to ensure the guidelines are accurately and consistently applied during item development and that the intended construct, cognitive complexity, and language demands specified in the grade-level standards have not been significantly altered.

Follow checklist for evaluating the linguistically modified items

For each item, verify that:

- The construct being tested has not changed.
- The cognitive complexity of the item is appropriate.
- The following elements in the linguistically modified item maximize English language learner students' linguistic access:
    - Context.

- Graphics.
- Vocabulary/wording.
- Sentence structure.
- Format/style.

Methods used to verify that the test item has been appropriately linguistically modified include:

- Expert verification (for example, by a technical advisory committee, content and bias review committee, or independent external reviewer) that the construct has not changed and that the cognitive complexity of the item is appropriate.

- Statistical analyses (for example, analysis of variance, differential item functioning analysis, or factor analysis).

- Cognitive interviews.

Table D1. Linguistic modification guidelines and strategies

Desirable characteristics	Notes on approaches and criteria
Item context	

• Familiar to students.
• No cultural or linguistic bias.
• Minimal construct (no irrelevant words or phrases).
	• The context situates the problem (and may include description of relationship or interaction between location and time).
	• In the body of the report, context is often described in relation to its complexity and as part of biased or construct-irrelevant information that should be pruned out. Recommendations:
	  o Remove passive voice construction in original item.
	  o Remove past tense and conditional in original item.
	  o Break stem into shorter, less complex sentences (sometimes a series of shorter sentences can create a story line or present a more familiar context/situation to students).
	• Context can provide description that helps make abstract or highly generalized situations more concrete and relevant. Simply stated, it helps to ground the content being tested. Context that facilitates access for English language learner students is expressed in concrete language, illustrative language, and illustrations/graphics.

Desirable characteristics	Notes on approaches and criteria
Item graphics	
• Familiar to students. • No cultural or linguistic bias. • Symbols, legends, and key vocabulary relevant to the construct and familiar to English language learner students. • Consistent graphic and labeling/naming conventions • Supportive of English language learner student understanding of assessed content.	• Graphics include diagrams, tables, charts, drawings, graphs, pictures, and maps. • Student knowledge about certain graphics is required and assessed in mathematics. • Graphics allow for reduced amount or complexity of language in a test item. Use of graphics in test items should serve a clear purpose. Otherwise they may be misleading or distracting. For example, graphics may be used to: o Clarify key aspects of the content/construct assessed. o Clarify construct-relevant context. o Clarify a mathematical operation. o Indicate what the student is expected to do. o Help students shift from one context to another within an assessment (for example, from one type of test item to another). o Allow students to reinforce or verify understanding of key information in test item. o Simplify the structure of a test item that requires a number of operations or steps (for example, through bulleted lists or a diagram of the complete problem that accurately reflects the problem in its totality). • Some criteria that can be used to evaluate the need for a graphic include: o Does the graphic clarify construct-irrelevant information? If so, it may not be necessary. It might be better to revise or delete the construct-irrelevant information. o Does the graphic support the test item context without requiring additional written text? o Does the graphic accurately represent the full complexity of the problem? If not, it may be misleading. o Is the graphic consistent with the key content/construct of the item?

Desirable characteristics	Notes on approaches and criteria
Item vocabulary/wording	
• High-frequency words. • Common and familiar words. • Relevant technical terms that reflect language of the content standards and academic English language. • Technical terms defined, as appropriate. • Naming conventions consistent with graphics/stimuli. • Construct-irrelevant vocabulary/phrases at or below grade level.	• Careful selection of vocabulary and phrases can simplify sentence structure. The amount and complexity of language should be balanced with the amount of information necessary for student to understand/access the item. The goal is to make the language as clear and straightforward as possible, while still providing the amount and complexity of information necessary to communicate the targeted content of the test item. • Some general guidelines: o Use precise language. Appropriate language modification does not simply mean using common or familiar vocabulary. o Consider language used in the content standards and academic English language . o Repeat key words/phrases in the test item that students need to understand the item and respond to it. o Do not automatically provide synonyms for a key word. This may not be helpful, especially if a test item is already long or complex. Although providing synonyms may be helpful during instruction, it may not be useful in assessment items. o Use words/phrases consistently within the context of the item and consider consistency of terms within a strand—for example, reading or measurement). Support this use with context-familiar content-based abbreviations and make explicit connections between terms/abbreviations. • If possible, avoid using: o Ambiguous words or unnecessary words with multiple meanings. o Irregularly spelled words. o Proper nouns that are irrelevant or not meaningful to the population. o Words that are both nouns and verbs (for example, carpet, value, cost); however, if a choice needs to be made, use the word only as a noun. o Hyphenated and compound words o Gerunds. o Relative pronouns (for example, which, who, that) without a clear antecedent.

Desirable characteristics	Notes on approaches and criteria
Item sentence structure	
• Familiar, common sentence structure. • Complexity of sentence structure at or below grade level. • Key information presented first or early in the test item. • One sentence per idea for complex test items.	• To reduce the complexity of a sentence in a test item: o Identify the agent (that is, the person or object carrying out the action) to construct sentences that use active voice (and avoid passive voice). o Make sure that the verb in a sentence follows the subject as closely as possible. o Remove introductory phrases that are irrelevant to the construct being tested. o Use conventional constructions (for example, apostrophes for possessives and "s" or "es" for plurals. o Use proper nouns that students are familiar and are grade-level appropriate. o Use clear grammatical structures. • To reduce language load: o Change past or future tense verb forms to present tense. o Change passive verb forms to active verb forms. o Change complex sentence structure to subject-verb-object structure. o Shorten any long nominals/names/phrases (for example, "last year's class vice-president" to "a student leader"). o Replace compound sentences with two separate sentences, especially when making comparisons. o Shorten or delete long prepositional phrases. o Replace conditional clauses with separate sentences. o Change the order of a clause within a sentence. o Remove or rephrase relative clauses. o Rephrase questions framed in negative terms. • Make sure the following are clear. o Noun-pronoun relationships. o Antecedent references.

Desirable characteristics	Notes on approaches and criteria
Item format/style	
• Clear parts of the item/question. • Explicit order of operations. • Relevant and appropriate distinctions. • Segmented or shortened long problem statements.	• Place test item elements in the following order: (1) text that introduces the graphic; (2) graphic; and (3) the test item stem. • Format for emphasis of key words/terms (highly construct-relevant), using bold, ALL CAPS, and underline to call English language learner students' attention to them. • Consider whether blocks of text (that is, a paragraph) may be necessary and appropriate for presenting a test item. This depends on the construct assessed, the complexity of the information needed by the student to respond to the item, and the centrality of the context to the construct. Suggested strategies to help English language learner students process such text include: o Bulleted lists. o Indenting key information. o Emphasizing key words/terms. o Using graphics.

Source: Sato 2008.

Key terms

This section described key terms used in the discussion of linguistically modified assessments for training item developers.

Access

To maximize student access to the content being assessed on an achievement test (for example, mathematics), text in the item that is not directly related to the targeted construct (that is, construct-irrelevant text) is minimized or removed. Doing so facilitates students' ability to demonstrate their construct-relevant knowledge and skills and reduces or eliminates sources of construct-irrelevant variance (construct irrelevance) in test results among students. In other words, when access is constrained, it can result in the measurement of sources of variance that are not related to the intended test content. If student access to tested content is restricted, students cannot fully demonstrate what they know and can do; subsequently, test results underestimate their level of content achievement (underrepresentation).

In this study the construct-irrelevant factors that constrain access to tested content for English language learner students are examined to support development of mathematics test items that maximize students' ability to show what they know and can do in mathematics.

Accommodation vs. modification

An accommodation is a change in testing conditions that is implemented to increase accessibility of test content to a specific student population. Such changes are deemed fair and reasonable when standardized administration conditions do not provide an equal opportunity for all students to demonstrate what they know and can do (Abedi & Lord 2001; Butler & Stevens 2001; Holmes & Duron 2000; National Research Council 2002, 2004). It is assumed that the same construct is being assessed with and without the accommodation. An accommodation is intended to minimize or remove the effects on test performance of construct-irrelevant factors that may contribute to, for example, the underrepresentation of student achievement in the content area.

A modification is an adjustment to the test itself, the administration conditions, or the content standards for assessment. While modification may improve access to the test content for a specific student population in a fair and reasonable manner, it significantly alters the construct being assessed. Examples of test modifications include allowing students with specific disabilities to use calculators on mathematics computation items (when general education students cannot) or allowing the reading comprehension portions of a test to be read aloud to English language learner students.

In traditional psychometric practice, accommodations may affect the performance of its intended referent group only, while remaining construct-neutral to nonaccommodated students—that is,

---

characteristics. However, evaluation can be done only at the discourse level. A critical reading and assignment of meaning requires minimum language beyond the word or sentence level.

the accommodation should benefit the student needing the accommodation but should have no effect on those not needing the accommodation.

However, research-based test design practices (for example, universal design, simplified language in items and associated text) suggest that all student groups may benefit from item development strategies designed to minimize construct-irrelevant variance. So, for this study an accommodation may be considered valid, even if all groups benefit from its use, if evidence collected suggests that:

- The construct/content assessed was not significantly altered.
- The performance of the group targeted for accommodation (that is, English language learner students) improves at a greater rate than that of their English-proficient counterparts.

English language learner students

English language learner students are "national-origin-minority students[39] who cannot speak, read, write, or comprehend English well enough to participate meaningfully in and benefit from the schools' regular education program" (U.S. Department of Education, Office of Elementary and Secondary Education 1999, p. 60). No Child Left Behind legislation (including Title III) refers to this population as "limited English proficient" (U.S. Department of Education, Office of Elementary and Secondary Education 2000).

This study's analyses included only students in grades 7 and 8 who identified themselves as "Hispanic" or who identified Spanish as their first language or the language spoken in their home. Recruitment efforts targeted Spanish-speaking English language learner students who scored at the mid- to high range of English language proficiency to ensure that their command of the English language was at a level sufficient to benefit from the linguistic modification.

Linguistic modification

Linguistic modification is a theory- and research-based process in which the language in test items, directions, and response options is modified in ways that clarify and simplify the text without simplifying or significantly altering the construct assessed. To facilitate comprehension, linguistic modification reduces construct-irrelevant language demands (for example, semantic and syntactic complexity) of text through strategies such as reduced sentence length and complexity, use of common or familiar words, and use of concrete language (Abedi et al. 2005; Abedi, Lord, & Plummer 1997; Sireci, Li, & Scarpati 2002).

Linguistic modification is not simply good editing practice and does not result in simpler items. Rather, it is a linguistically based, systematic means for targeting, reducing, and removing the irrelevant variance in test performance that is attributable to individual differences in English proficiency so that English language learner students can fully demonstrate what they know and

[39] "National origin minority" can include students born in the United States.

can do in that content area. By minimizing the language load, a source of construct-irrelevant variance, English language learner students' access to construct-relevant content is enhanced.

Appendix N. Operational item set—original

Research Study

OPERATIONAL TEST FORM-O

Math Test

Grades 7&8

2008

Student Name:

_____

REL WEST

3. Fifteen boxes each containing 8 radios can be repacked in 10 larger boxes each containing how many radios?

A. 8

B. 12

C. 80

D. 120

7. What is 4 hundredths written in decimal notation?

A. 0.004

B. 0.04

C. 0.400

D. 4.00

10. If Jill is driving at 65 miles per hour, what is her approximate speed in kilometers per hour? (1 mile ≈ 1.6 kilometers)

A. 16

B. 41

C. 104

D. 173

11. A certain reference title contains approximately one billion facts. About how many millions is that?

A. 1,000,000

B. 100,000

C. 10,000

D. 1,000

12. A car odometer registered 41,256.9 miles when a highway sign warned of a detour 1,200 feet ahead. What will the odometer read when the car reaches the detour? (5,280 feet = 1 mile)

A. 42,456.9

B. 41,261.3

C. 41,259.2

D. 41,257.1

14. The mean distance from Venus to the Sun is 1.08×10^8 kilometers. Which of the following quantities is equal to this distance?

A. 10,800,000 kilometers

B. 108,000,000 kilometers

C. 1,080,000,000 kilometers

D. 10,800,000,000 kilometers

15. If the values of the expressions below are plotted on a number line, which expression would be closest to five?

A. $|-4|$

B. $|-18|$

C. $|7|$

D. $|16|$

17. A sweater originally cost $37.50. Last week, Moesha bought it at 20% off.

How much was deducted from the original price?

A. $7.50

B. $17.50

C. $20.00

D. $30.00

20. A landscaper estimates that landscaping a new park will take 1 person 48 hours. If 4 people work on the job and they each work 6-hour days, how many days are needed to complete the job?

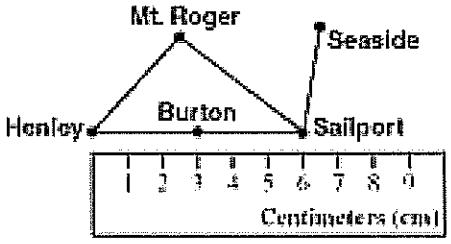
A. 2 days

B. 4 days

C. 6 days

D. 8 days

24. Javier is using a ruler and a map to measure the distance from Henley to Sailport.



The actual distance from Henley to Sailport is 120 kilometers (km). What scale was used to create the map?

A. 1 cm = 6 km

B. 1 cm = 12 km

C. 1 cm = 15 km

D. 1 cm = 20 km

Appendix O. Operational item set—linguistically modified

Research Study

OPERATIONAL TEST FORM-M

Math Test

Grades 7&8

2008

Student Name: _____

REL WEST

3. A student works in a store.

- She unpacks 15 boxes.
- Each box contains 8 radios.
- She repacks the radios in 10 larger boxes.
- Each box contains the same number of radios.

How many radios are in each larger box?

A. 8

B. 12

C. 80

D. 120

7. 4 hundredths = _____

A. 0.004

B. 0.04

C. 0.400

D. 4.00

10. 65 miles per hour is about _____ kilometers per hour
(1 mile ≈ 1.6 kilometers)

A. 16

B. 41

C. 104

D. 173

11. How many millions is 1 billion?

A. 1,000,000

B. 100,000

C. 10,000

D. 1,000

12. A car's mileage is 41,256.9 miles. The car travels 1,200 feet to an exit. What is the car's mileage at the exit? (5,280 feet = 1 mile)

A. 42,456.9

B. 41,261.3

C. 41,259.2

D. 41,257.1

14. Which distance equals 1.08×10^8 kilometers?

A. 10,800,000 kilometers

B. 108,000,000 kilometers

C. 1,080,000,000 kilometers

D. 10,800,000,000 kilometers

15. Which value is closest to five on a number line?

A. $|-4|$

B. $|-16|$

C. $|7|$

D. $|16|$

17. A girl wants to buy a sweater on sale.

- The regular price is $37.50.
- The discount is 20% of the regular price.

What is the amount of the discount?

A. $7.50

B. $17.50

C. $20.00

D. $30.00

20. A manager hires students to do a job.

- She estimates that 1 student needs 48 hours to do the job.
- She hires 4 students to do the job together.
- Each student works 6 hours per day.

What is the total number of days the 4 students need to do the job?

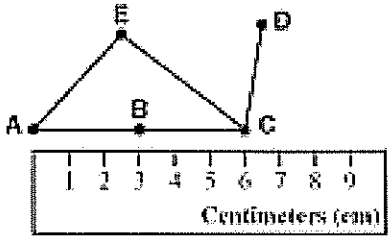
A. 2 days

B. 4 days

C. 6 days

D. 8 days

24. Look at the map and ruler below. The diagram below shows the distance from Point A to Point C on a map.



The actual distance from Point A to Point C is 120 kilometers (km). What is the scale of the map?

A. 1 cm = 6 km

B. 1 cm = 12 km

C. 1 cm = 15 km

D. 1 cm = 20 km

Item Number: _____

Level of Cognitive Complexity	Language that <u>should not</u> be simplified or changed	Language that can/should be simplified or changed

Evaluation of Item Elements for Plain English: Accessibility of Content		
Item Context	**Item Graphics**	**Item Vocabulary/ Wording**

Evaluation of Item Elements for Plain English: Accessibility of Content		
Item Sentence Structure	**Item Format/ Style**	**Other/Comments**

Revised Item:

19. When he left the pizza restaurant, Joseph had 25 pizzas to deliver. At his first stop, he delivered five pizzas to a party. At his second stop, he delivered half of the remaining pizzas to a school. At each remaining stop, he delivered one pizza. How many stops did Joseph make to deliver the 25 pizzas?

A 3

B 10

C 12

D 25

20. Morgan's family made a large pizza for lunch on Saturday. Morgan ate $\frac{3}{12}$ of the pizza. Megan ate $\frac{1}{6}$ of the pizza, and Emma ate $\frac{1}{12}$ of the pizza. Their parents ate $\frac{1}{3}$ of the pizza. How much pizza was left?

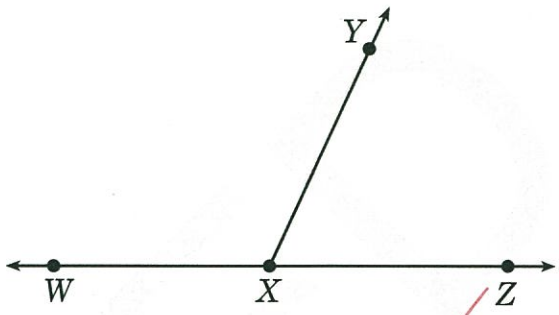
A $\frac{1}{12}$

B $\frac{1}{6}$

C $\frac{6}{12}$

D $\frac{5}{6}$

21. ***About*** how many degrees is the measure of $\angle WXY$?



A 20°

B 60°

C 120°

D 160°

22. Joey was looking at a square, a rectangle, and a right triangle. What is the total number of angles for all of the polygons, and how many are right angles?

A 11 angles, 8 right angles

B 11 angles, 9 right angles

C 12 angles, 8 right angles

D 12 angles, 9 right angles

19. Cara used this multiplication table to help her find the quotient for $112 \div 14$.

Multiplication Table

×	10	11	12	13	14	15	16
6	60	66	72	78	84	90	96
7	70	77	84	91	98	105	112
8	80	88	96	104	112	120	128
9	90	99	108	117	126	135	144
10	100	110	120	130	140	150	160
11	110	121	132	143	154	165	176

What answer should Cara get?

A 16

B 11

C 8

D 7

20. Mrs. Jones has some baskets of strawberries to sell. She has 52 baskets each containing 3 pounds of strawberries and 48 smaller baskets each containing 2 pounds of strawberries. ***About*** how much will her strawberries weigh in all?

A 250 pounds

B 200 pounds

C 150 pounds

D 100 pounds

21. Sallie baked 4 apple pies and cut each of them into sixths. If she served $3\frac{1}{2}$ pies, how many slices of pie did Sallie serve?

A 24

B 21

C 18

D 9

22. Clint's teacher asked him to write two fractions that are equivalent to $\frac{2}{5}$. If Clint did this problem correctly, which answer did Clint write?

A $\frac{2}{10}$ and $\frac{4}{10}$

B $\frac{4}{10}$ and $\frac{6}{10}$

C $\frac{2}{10}$ and $\frac{20}{100}$

D $\frac{4}{10}$ and $\frac{40}{100}$

16. Which chart shows the rule that the output value is two less than the input value?

A

Input	Output
5	7
8	10
11	13
12	14

B

Input	Output
5	3
8	4
11	9
12	10

C

Input	Output
5	10
8	16
11	22
12	24

D

Input	Output
5	3
8	6
11	9
12	10

17. The bread truck makes deliveries to a store 3 days each week. Each delivery has 45 loaves of bread. Which expression could be used to determine the number of loaves of bread delivered in 5 weeks?

A 3×5

B $45 \div (3 \times 5)$

C 45×3

D $45 \times 3 \times 5$

18. Michael cuts grass for $15.00 per lawn. He cuts 2 lawns each day for 6 days a week. How much will Michael earn in 2 weeks?

A $390

B $360

C $180

D $90

1. The library has 7,126 books. The library will purchase exactly one hundred more books. How many books will the library have after the books are purchased?

A 7,136

B 7,137

C 7,226

D 8,126

2. There are 20 seeds in a package. If 5 seeds are put in each flower pot, how many flower pots are needed to plant all of the seeds?

A 4

B 5

C 15

D 25

3. A box of candy has 12 rows. There are 6 pieces of candy in each row. How many pieces of candy are in the box?

A 6

B 18

C 62

D 72

4. On Saturday, 2,759 people went to the afternoon concert and 6,387 people went to the night concert. ***About*** how many people went to the concert on Saturday?

A 4,000

B 6,000

C 8,000

D 9,000

5. Dean had 1,062 pennies in his bank. Shawn had 889. How many more pennies did Dean have than Shawn?

A 173

B 223

C 227

D 283

6. Jerry keeps his rock collection in 7 boxes. Each box weighs about 6 or 7 pounds. How much does Jerry's whole rock collection weigh?

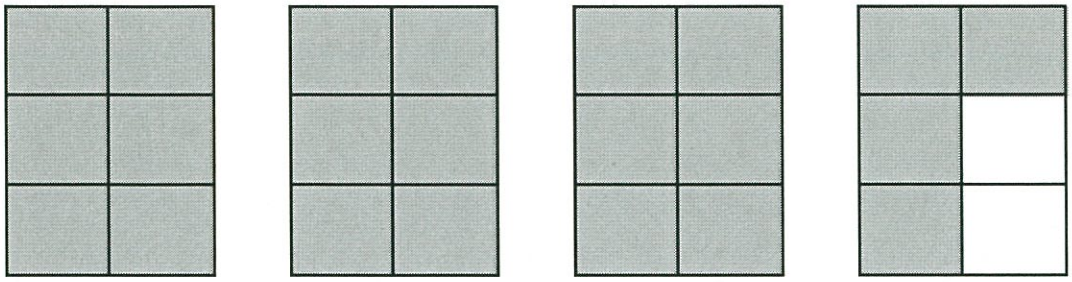
A between 50 and 60 pounds

B between 40 and 50 pounds

C between 30 and 40 pounds

D between 20 and 30 pounds

1. Which mixed number represents the shaded parts of the model?



A $3\frac{2}{6}$

B $3\frac{4}{6}$

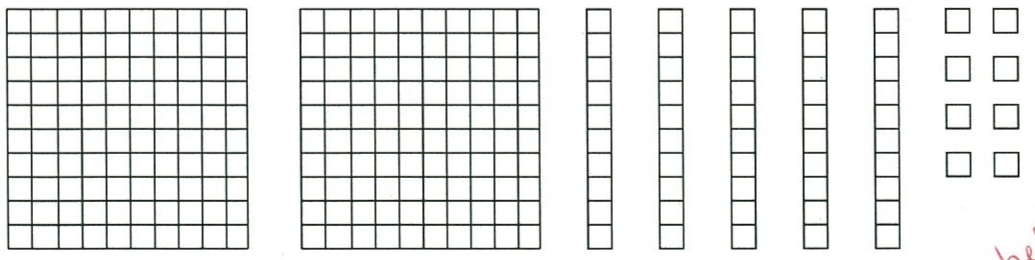
C $4\frac{2}{6}$

D $4\frac{4}{6}$

- Introduce Model or label it.
- Q. Is model part of curriculum
- add "shown below" at end
- another word for "model"? Boxes Cells

---

2. Which number is 100 more than the model shown below?



A 158

B 258

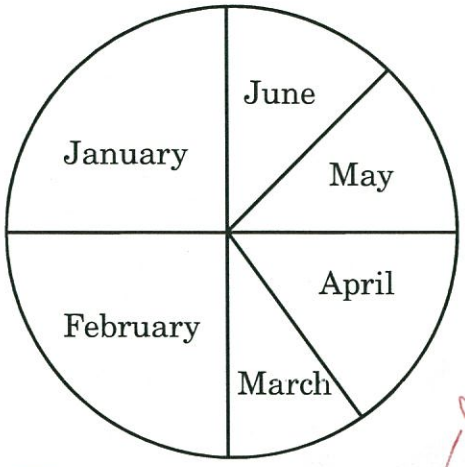
C 358

D 385

Is there better placement for the stem?

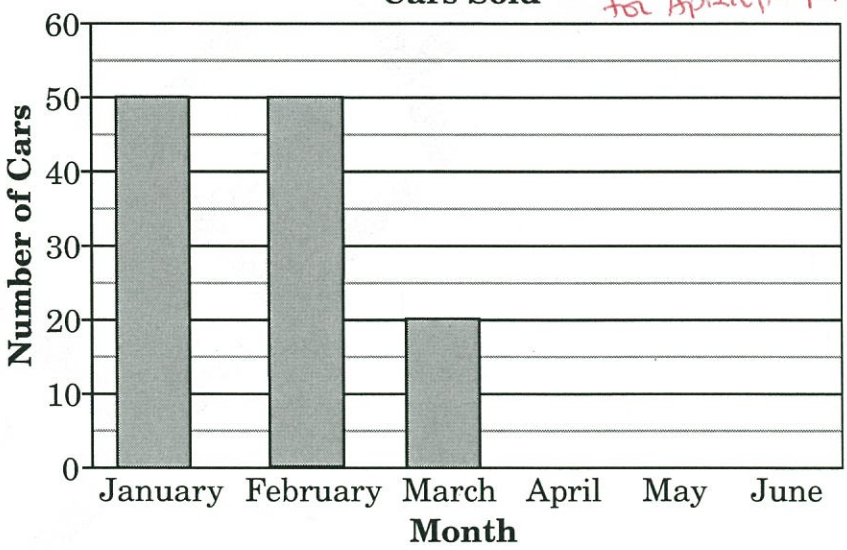
30. A dealership sold 200 cars in a six-month period. The circle graph below displays the distribution of sales by month.

Distribution of Car Sales



The sales manager at the dealership created the bar graph below to show the number of cars sold each month during the six-month period. The bars for April, May, and June have not yet been drawn.

Cars Sold



The dealership sold the same number of cars in June as in May. How many cars did it sell in April?

A 20

B 25

C 30

D 35

Appendix 3-C
*NCEXTEND*1 Sample Item Writer Training PPT

Item Critique

Thank you for your interest in writing ***NCEXTEND*1** questions for the Grade 10 reading test.

The following presentation will discuss items that need revisions and possible ways to revise them.

Characteristics of a Quality ***NCEXTEND1*** Reading Item

- The item matches only one content goal.
- The item is written for students being instructed in the Extended Essential Standards for Grade 10 English Language Arts.
- The item has only one best answer.
- The item has three foils.

Characteristics of a Quality ***NCEXTEND1*** Reading Item

The item is written using the format below.

- Present:
  - Stimulus card: (if stimulus card is used)
- **SAY:** State what the assessor is to say to the student in relation to the stimulus card (if stimulus card is used).
- **SAY:** Ask the student the item question.
- Present cards in the following order: (three foils)
  - Card A:
  - Card B:
  - Card C:
- **SAY:** This says _____. This says _____. This says _____. State the content of each card. (If the content gives the answer, do not state any of the foils.)
- **SAY:** Repeat the same item question asked above. State the question again in statement form, using **Show me …**, as the beginning of the statement.

Characteristics of a Quality ***NCEXTEND1*** Reading Item

- The item does not use “not” in the stem.
- A question is always asked before foils are shown.
- The stem of the question is written with few extraneous words. Avoid “teaching” in the stem.
- The foils are parallel.
- Foils are each close to the same length or ordered from shortest to longest.
- The exact question asked before showing the foils is repeated after foils are shown and described.
- Foils are not described if the description would give the answer to the question.
- The final statement to the student begins with the words, “Show me.”

Next, you will see some sample items that need improving. See if you can spot problems with them.

Sample Item

- **SAY:** Why does Alice get a big box?
- Present cards in the following order:
  - Card A: To hide the cat.
  - Card B: To use as a table.
  - Card C: To mail the books.
- **SAY:** Why does Alice get a big box? Show me why Alice gets a big box.

Sample Item

- **SAY:** Why does Alice get a big box?
- Present cards in the following order:
  - Card A: To hide the cat.
  - Card B: To use as a table.
  - Card C: To mail the books.
- **SAY:** Why does Alice get a big box? Show me why Alice gets a big box.

The foils have not been described to the student.

Sample Item
(Corrected)

- **SAY:** Why does Alice get a big box?
- Present cards in the following order:
  - Card A: To hide the cat.
  - Card B: To use as a table.
  - Card C: To mail the books.
- **SAY:** This says to hide the cat. This says to use as a table. This says to mail the books.
- **SAY:** Why does Alice get a big box? Show me why Alice gets a big box.

Sample Item

- Present cards in the following order:
  - Card A: Maria
  - Card B: Von
  - Card C: Leo
- **SAY:** This says Maria. This says Von. This says Leo.
- **SAY:** Who is the new classmate Mike is talking about? Show me the new classmate Mike is talking about.

Sample Item

- Present cards in the following order:
  - Card A: Maria
  - Card B: Von
  - Card C: Leo
- **SAY:** This says Maria. This says Von. This says Leo.
- **SAY:** Who is the new classmate Mike is talking about? Show me the new classmate Mike is talking about.

A question is not asked before the foils are presented.

Sample Item Corrected

- **SAY:** Who is the new classmate Mike is talking about?
- Present cards in the following order:
  - Card A: Maria
  - Card B: Von
  - Card C: Leo
- **SAY:** This says Maria. This says Von. This says Leo.
- **SAY:** Who is the new classmate Mike is talking about? Show me the new classmate Mike is talking about.

Sample Item

- **SAY:** What else might the class put in the new library?
- Present cards in the following order:
  - Card A: student drawings
  - Card B: dance floor
  - Card C: snack bar
- **SAY:** This says student drawings. This says dance floor. This says snack bar.
- **SAY:** What else might the class put in the new library?

Sample Item

- **SAY:** What else might the class put in the new library?
- Present cards in the following order:
  - Card A: student drawings
  - Card B: dance floor
  - Card C: snack bar
- **SAY:** This says student drawings. This says dance floor. This says snack bar.
- **SAY:** What else might the class put in the new library?

There is no final “Show me…” statement.

Sample Item (Corrected)

- **SAY:** What else might the class put in the new library?
- Present cards in the following order:
  - Card A: student drawings
  - Card B: dance floor
  - Card C: snack bar
- **SAY:** This says student drawings. This says dance floor. This says snack bar.
- **SAY:** What else might the class put in the new library? Show me what else the class might put in the new library.

Sample Item

- **SAY:** Mr. Martin brought an origami crane to class. Which of these is an origami crane?
- Present cards in the following order:
  - Card A: hawk
  - Card B: eagle
  - Card C: crane
- **SAY:** This says hawk. This says eagle. This says crane.
- **SAY:** Which of these is an origami crane? Show me an origami crane.

Sample Item

- **SAY:** Mr. Martin brought an origami crane to class. Which of these is an origami crane?
- Present cards in the following order:
  - Card A: hawk
  - Card B: eagle
  - Card C: crane
- **SAY:** This says hawk. This says eagle. This says crane.
- **SAY:** Which of these is an origami crane? Show me an origami crane.

The foil descriptions give the answer.

No foil descriptions should occur in this item.

Sample Item (Corrected)

- **SAY:** Mr. Martin brought an origami crane to class. Which of these is an origami crane?
- Present cards in the following order:
  - Card A: hawk
  - Card B: eagle
  - Card C: crane
- **SAY:** Which of these is an origami crane? Show me an origami crane.

Sample Item

- **SAY:** Many fish live in coral reefs. The are a food source for other fish and an important part of the ocean food chain. What does the author feel is the problem with coral reefs?
- Present cards in the following order:
  - Card A: They are being polluted.
  - Card B: They are growing too fast.
  - Card C: They are hiding fish from fishermen.
- **SAY:** This says they are being polluted. This says they are growing too fast. This says they are hiding fish from fishermen. What does the author feel is the problem with coral reefs? Show me what the author feels is the problem with coral reefs.

Sample Item

- **SAY:** Many fish live in coral reefs. The are a food source for other fish and an important part of the ocean food chain. What does the author feel is the problem with coral reefs?
- Present cards in the following order:
  - Card A: They are being polluted.
  - Card B: They are growing too fast.
  - Card C: They are hiding fish from fishermen.
- **SAY:** This says they are being polluted. This says they are growing too fast. This says they are hiding fish from fishermen. What does the author feel is the problem with coral reefs? Show me what the author feels is the problem with coral reefs.

There is too much teaching in the stem.

Sample Item (Corrected)

- **SAY:** What does the author feel is the problem with coral reefs?
- Present cards in the following order:
  - Card A: They are being polluted.
  - Card B: They are growing too fast.
  - Card C: They are hiding fish from fishermen.
- **SAY:** This says they are being polluted. This says they are growing too fast. This says they are hiding fish from fishermen. What does the author feel is the problem with coral reefs? Show me what the author feels is the problem with coral reefs.

Sample Item

- **SAY:** What is a way to conserve water?
- Present cards in the following order:
  - Card A: Collect rainwater for plants.
  - Card B: Serve water in all restaurants.
  - Card C: Post pictures about drought.
- **SAY:** This says collect rainwater for plants. This says serve water in all restaurants. This says draw pictures about drought.
- What is a way to conserve water? Show me a way to conserve water.

Sample Item

- **SAY:** What is a way to conserve water?
- Present cards in the following order:
  - Card A: Collect rainwater for plants.
  - Card B: Serve water in all restaurants.
  - Card C: Post pictures about drought.
- **SAY:** This says collect rainwater for plants. This says serve water in all restaurants. This says draw pictures about drought.
- What is a way to conserve water? Show me a way to conserve water.

There are two correct answers. Both A and C could result in conservation of water.

Sample Item (Corrected)

- **SAY:** What is a way to conserve water?
- Present cards in the following order:
  - Card A: Collect rainwater for plants.
  - Card B: Serve water in all restaurants.
  - Card C: Play in a sprinkler.
- **SAY:** This says collect rainwater for plants. This says serve water in all restaurants. This says play in a sprinkler. What is a way to conserve water? Show me a way to conserve water.

Item Critique: Manipulatives

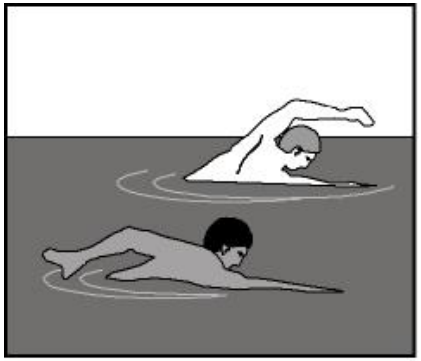
Item Critique: Manipulatives

- When reviewing items it is important to carefully examine the manipulatives that accompany the item.
- In each of the next items there is something wrong with the item's manipulatives. Each error will be pointed out.

Item

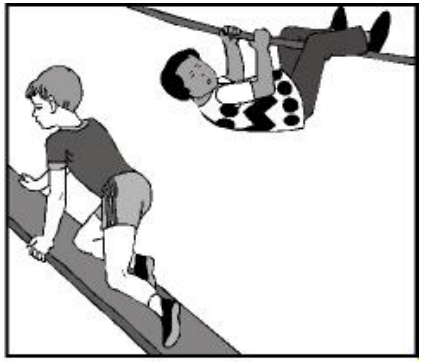
- **SAY:** How are Nick and Sam alike?
- Present cards in the following order:
  - Card A: Both like to swim.
  - Card B: Both like to read.
  - Card C: Both like to climb.
- **SAY:** This says both liked to swim. This says both liked to read. This says both liked to climb.
- **SAY:** How are Nick and same alike? Show me how Nick and Sam are alike.

Item Manipulatives



Both like to swim

Both like to read



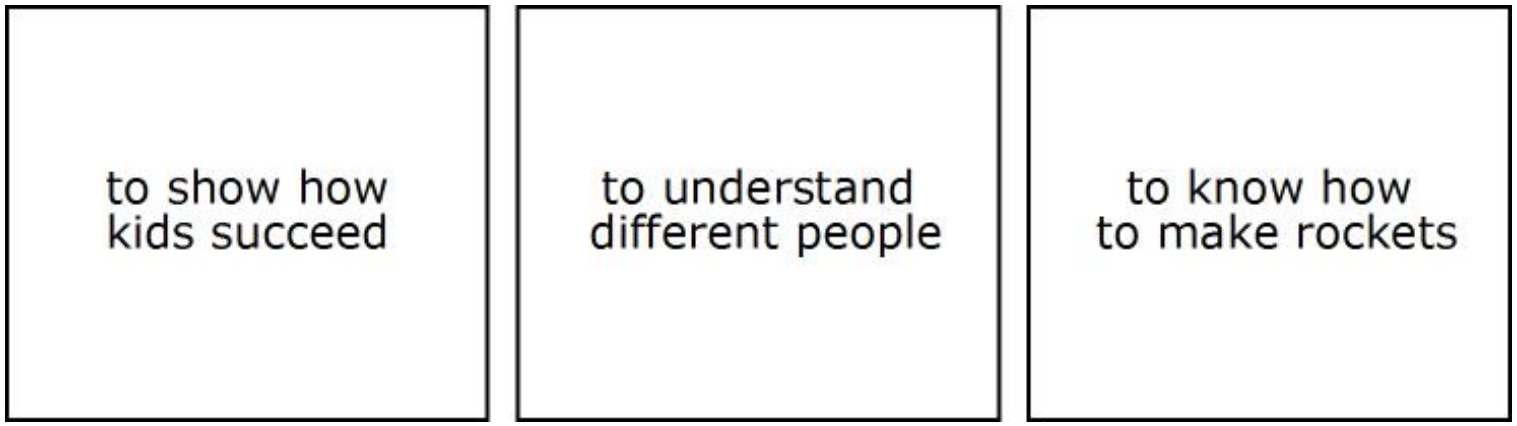
Both like to climb

- In the selection Nick is helping Sam learn to swim. Nick is in high school and Sam is in grade 2. The boys in the manipulatives are too similar in age. The depictions do not match the story.

Item

- **SAY:** Why did Matt want the class to read *One Bright Star*?
- Present cards in the following order:
  - Card A: to show how kids succeed
  - Card B: to understand different people
  - Card C: to know how to make rockets
- **SAY:** This says to show how kids succeed. This says to understand different people. This says to know how to make rockets.
- **SAY:** Why did Matt want the class to read *One Bright Star*? Show me why Matt want the class to read *One Bright Star*.

Item Manipulatives

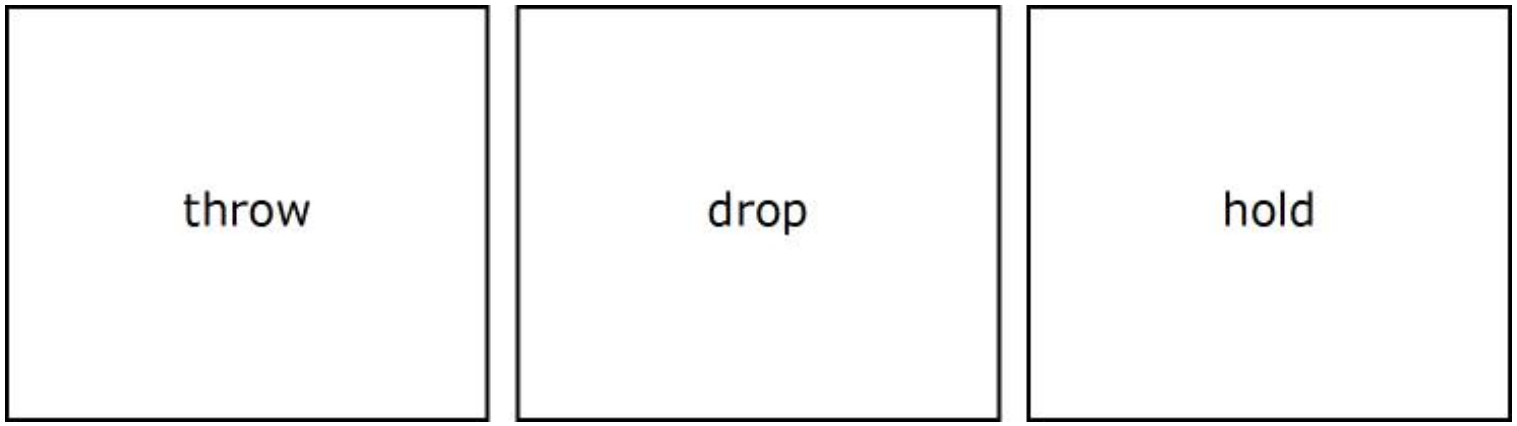


- In the selection Matt was talking to the students about careers. He did want kids to succeed and to understand different people but he was really focused on jobs in the space industry. The manipulatives did not include a choice that indicated he wanted students to consider the space industry as a career.

Item

- **SAY:** Emily's goal is to be a doctor. What does goal mean?
- Present cards in the following order:
  - Card A: hope
  - Card B: problem
  - Card C: class
- **SAY:** This says hope. This says problem. This says class.
- **SAY:** What does goal mean? Show me what goal means.

Item Manipulatives



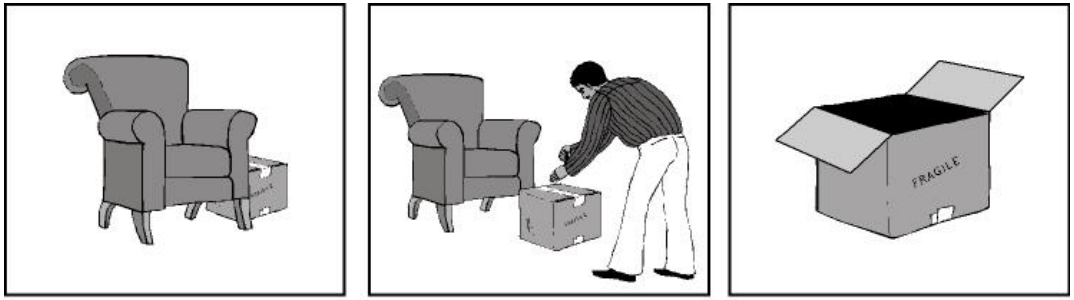
- The manipulatives do not match the item.

Item

- Read the story to the student.
- The story may be read as many times as necessary. The student must read the item independently and indicate his/her answer.
- **SAY:** Read the question and show me your answer.
- Present cards in the following order:
  - Question Card: What did Marty do first?
  - Card A: hide the box
  - Card B: find the box
  - Card C: open the box

Item Manipulatives

"What did Marty do first?"



Hide the box | Find the box | Open the box

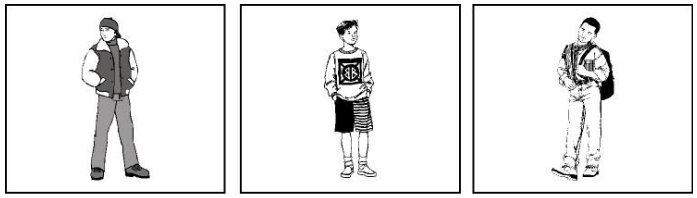
- In the selection Marty hid the box in the closet. The manipulatives do not depict the actions of the selection.

Item

- Present the story to the student.
- **SAY:** Read this story.
- The story may be read as many times as necessary. The student must read the item independently and indicate his/her answer.
- **SAY:** Read the question and show me your answer.
- Present cards in the following order:
  - Question Card: Who did Jon want to be friends with?
  - Card A: Alan
  - Card B: Jeff
  - Card C: Burt

Item Manipulatives

"Who did Jon want to be friends with?"



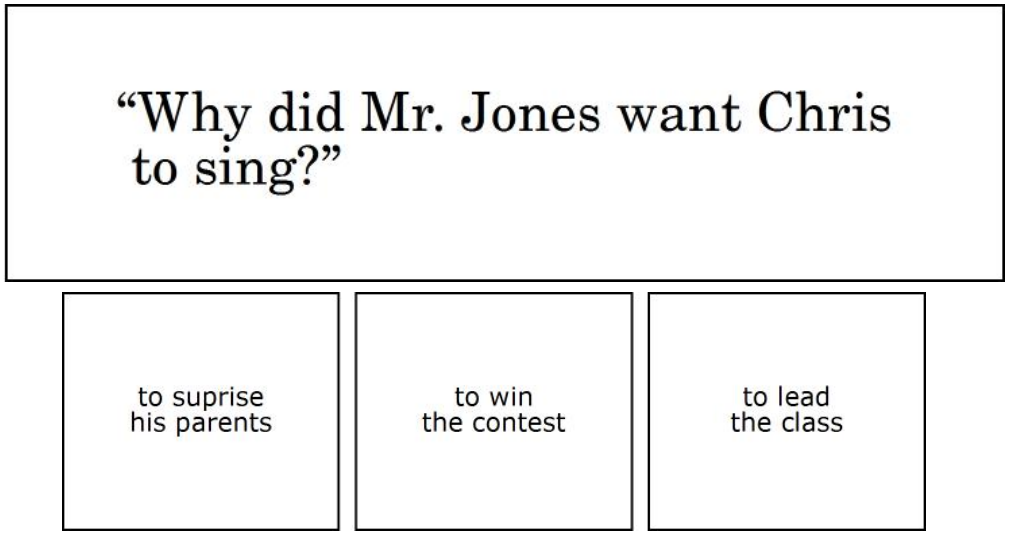
Alan Jeff Burt

- In the selection Alan looks like Jeff. The names are under the wrong pictures.

Item

- Present the story to the student.
- **SAY:** Read this story.
- The story may be read as many times as necessary. The student must read the item independently and indicate his/her answer.
- **SAY:** Read the question and show me your answer.
- Present cards in the following order:
  - Question Card: Why did Mrs. Jones want Chris to sing?
  - Card A: to surprise his parents
  - Card B: to win the contest
  - Card C: to lead the class

Item Manipulatives



- The question should read "Mrs. Jones" not "Mr. Jones."

Appendix 3-D
*NCEXTEND*1 Sample Items

Released Form

NCEXTEND1
English Language Arts/ Reading Assessment— Grade 4

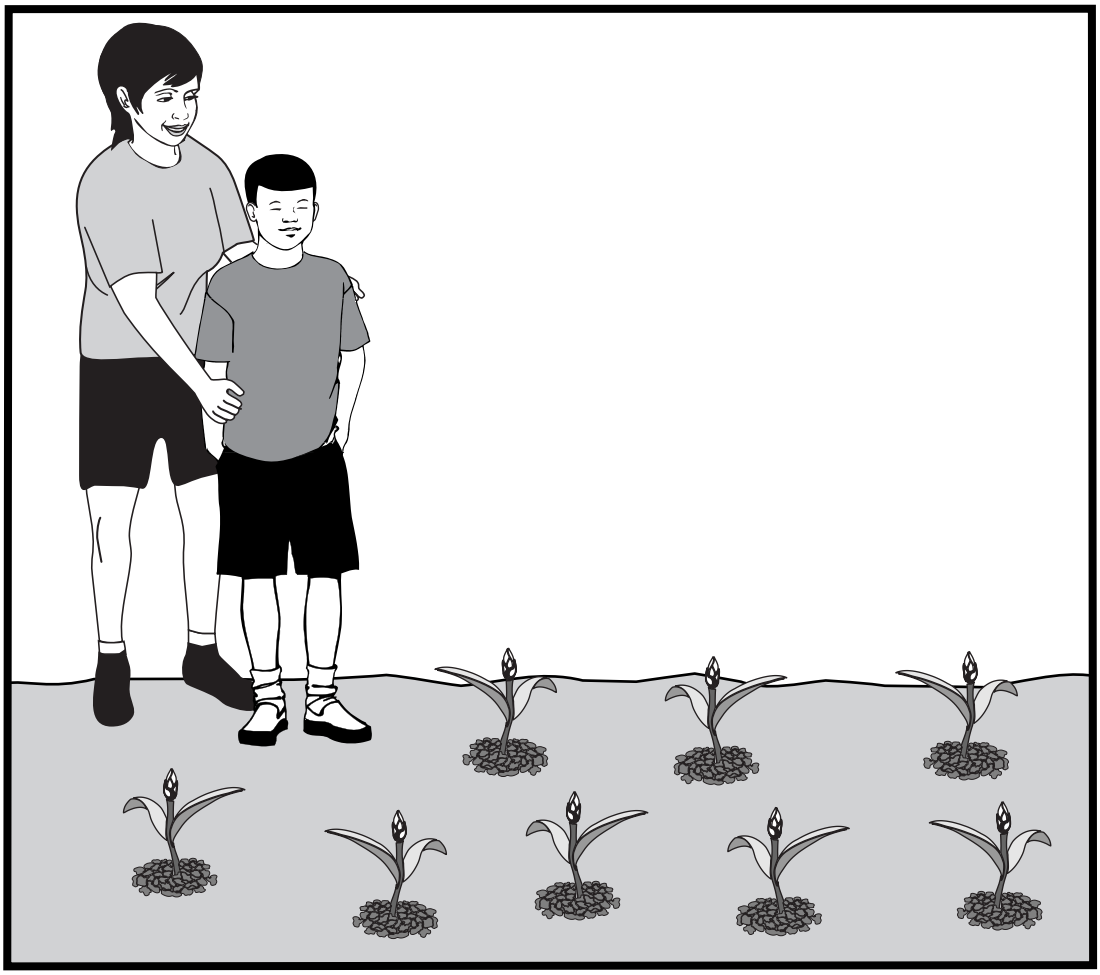
Assessor Booklet

Academic Services and Instructional Support
Division of Accountability Services

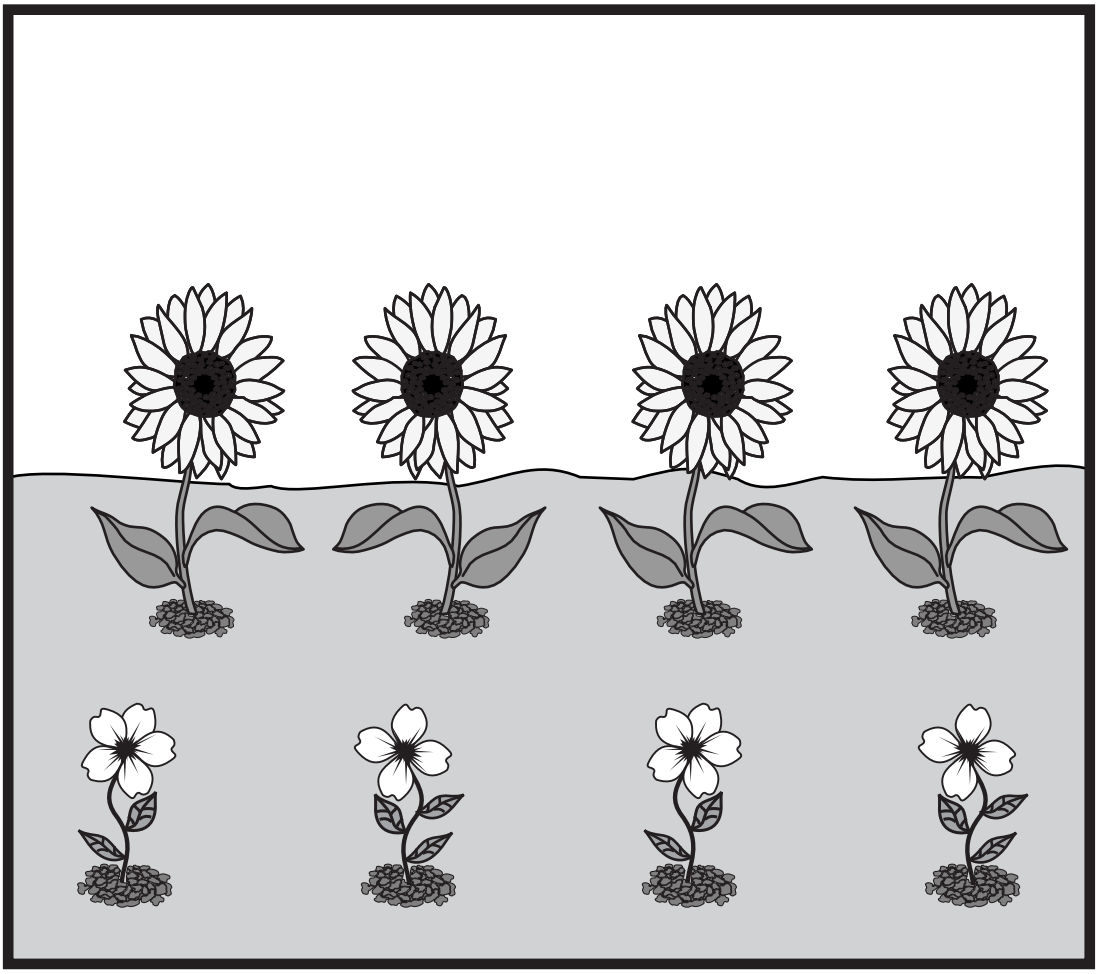
The Garden

Rafael and his mother are planting a garden. First, they plant seeds.

Then, they water the seeds.



They wait for the seeds to grow.



The seeds grow into big flowers and small flowers.

Rafael and his mother put some flowers in a vase. They like to look at the flowers.

Item 1

Manipulatives: Provided by NCDPI

- Selection: The Garden
- Card A: They plant seeds.
- Card B: They water the seeds.
- Card C: They wait for the seeds to grow.

*Objects/symbols may be substituted for the pictures if used routinely in the classroom. (Provided by the assessor)

Trial 1

- SAY: **"What is the first thing Rafael and his mother do to start their garden?"**
- Present the response cards in the following order using the script below (*Card A*, *Card B*, *Card C*).
- SAY: (Card A) **"They plant seeds."** (Card B) **"They water the seeds."** (Card C) **"They wait for the seeds to grow."**
- SAY: **"What is the first thing Rafael and his mother do to start their garden? Choose a card."**
- Record the student response below and follow the directions to continue with Trial 2 or the next item.

Fill in Trial 1 Student Response Here:	
Card A ⬭ →	Go to the next item.
Card B ⬭ →	Remove Card B, Trial 2
Card C ⬭ →	Remove Card C, Trial 2
No Response ⬭ →	Remove Card B, Trial 2

Trial 2

- SAY: **"What is the first thing Rafael and his mother do to start their garden?"**
- Present the response cards using the following script.
  If Card B was removed
- SAY: (Card A) **"They plant seeds."** (Card C) **"They wait for the seeds to grow."**
  If Card C was removed
- SAY: (Card A) **"They plant seeds."** (Card B) **"They water the seeds."**
- SAY: **"What is the first thing Rafael and his mother do to start their garden? Choose a card."**
- Record the student response below and continue to the next item.

Fill in Trial 2 Student Response Here:	
Card A	⬭ → Go to the next item.
Card B	⬭ → Go to the next item.
Card C	⬭ → Go to the next item.
No Response	⬭ → Go to the next item.

Item 2

Manipulatives: Provided by NCDPI

- Selection: The Garden
- Card A: The Seeds
- Card B: The Flowers
- Card C: The Watering Can

*Objects/symbols may be substituted for the pictures if used routinely in the classroom. (Provided by the assessor)

Trial 1

- SAY: **"What could be another title for this story?"**
- Present the response cards in the following order using the script below (*Card A*, *Card B*, *Card C*).
- SAY: (Card A) **"The Seeds"** (Card B) **"The Flowers"** (Card C) **"The Watering Can"**
- SAY: **"What could be another title for this story? Choose a card."**
- Record the student response below and follow the directions to continue with Trial 2 or the next item.

Fill in Trial 1 Student Response Here:	
Card A ⬭ →	Remove Card A, Trial 2
Card B ⬭ →	Go to the next item.
Card C ⬭ →	Remove Card C, Trial 2
No Response ⬭ →	Remove Card A, Trial 2

Trial 2

- SAY: **"What could be another title for this story?"**
- Present the response cards using the following script.
  If Card A was removed
- SAY: (Card B) **"The Flowers"** (Card C) **"The Watering Can"**
  If Card C was removed
- SAY: (Card A) **"The Seeds"** (Card B) **"The Flowers"**
- SAY: **"What could be another title for this story? Choose a card."**
- Record the student response below and continue to the next item.

Fill in Trial 2 Student Response Here:	
Card A	⬭ → Go to the next item.
Card B	⬭ → Go to the next item.
Card C	⬭ → Go to the next item.
No Response	⬭ → Go to the next item.

Item 3

Manipulatives: Provided by NCDPI

- Selection: The Garden
- Card A: picture of Rafael and his mother watering the seeds
- Card B: picture of Rafael and his mother waiting for the seeds to grow
- Card C: picture of Rafael and his mother looking at the flowers

*Objects/symbols may be substituted for the pictures if used routinely in the classroom. (Provided by the assessor)

Trial 1

- SAY: **"Which picture shows Rafael and his mother waiting for the seeds to grow?"**
- Present the response cards in the following order (*Card A*, *Card B*, *Card C*).
- SAY: **"Which picture shows Rafael and his mother waiting for the seeds to grow? Choose a card."**
- Record the student response below and follow the directions to continue with Trial 2 or the next item.

Fill in Trial 1 Student Response Here:	
Card A ⬭ →	Remove Card A, Trial 2
Card B ⬭ →	Go to the next item.
Card C ⬭ →	Remove Card C, Trial 2
No Response ⬭ →	Remove Card A, Trial 2

Trial 2

- SAY: **"Which picture shows Rafael and his mother waiting for the seeds to grow?"**
- Present the response cards in the following order.
  If Card A was removed, present Card B, then Card C
  If Card C was removed, present Card A, then Card B
- SAY: **"Which picture shows Rafael and his mother waiting for the seeds to grow? Choose a card."**
- Record the student response below and continue to the next item.

Fill in Trial 2 Student Response Here:	
Card A	⬭ → Go to the next item.
Card B	⬭ → Go to the next item.
Card C	⬭ → Go to the next item.
No Response	⬭ → Go to the next item.

Item 4

Manipulatives: Provided by NCDPI

- Selection: The Garden
- Stimulus: Rafael and his mother put some flowers in a _____.
- Card A: garden
- Card B: pot
- Card C: vase

*Objects/symbols may be substituted for the pictures if used routinely in the classroom. (Provided by the assessor)

Trial 1

- Present the stimulus card using the following script.
- SAY: **"This says Rafael and his mother put some flowers in a blank. Which word completes the sentence from the story?"**
- Present the response cards in the following order using the script below (*Card A*, *Card B*, *Card C*).
- SAY: (Card A) **"garden"** (Card B) **"pot"** (Card C) **"vase"**
- SAY: **"Which word completes the sentence from the story? Choose a card."**
- Record the student response below and follow the directions to continue with Trial 2 and end the sample items.

Fill in Trial 1 Student Response Here:		
Card A	⬭ →	Remove Card A, Trial 2
Card B	⬭ →	Remove Card B, Trial 2
Card C	⬭ →	Turn the page to end the sample items.
No Response	⬭ →	Remove Card A, Trial 2

Trial 2

- Present the stimulus card using the following script.
- SAY: **"This says Rafael and his mother put some flowers in a blank. Which word completes the sentence from the story?"**
- Present the response cards using the following script.
  If Card A was removed
- SAY: (Card B) **"pot"** (Card C) **"vase"**
  If Card B was removed
- SAY: (Card A) **"garden"** (Card C) **"vase"**
- SAY: **"Which word completes the sentence from the story? Choose a card."**
- Record the student response below and end the sample items.

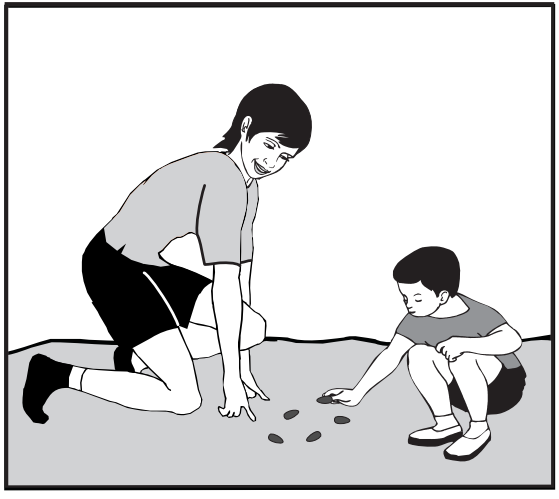
Fill in Trial 2 Student Response Here:	
Card A	⬭ → End the sample items.
Card B	⬭ → End the sample items.
Card C	⬭ → End the sample items.
No Response	⬭ → End the sample items.

End of the Sample Items

Read the following to announce the end of the sample items.

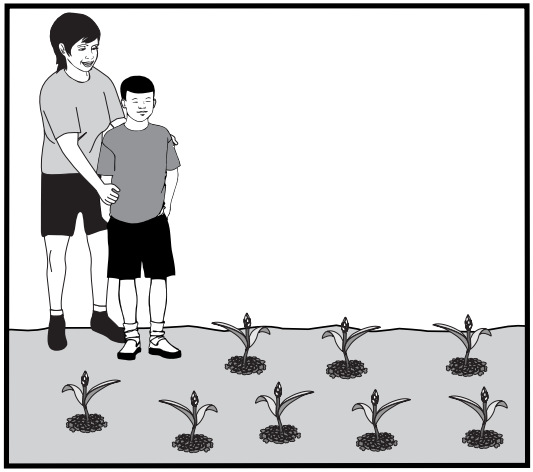
SAY: **"You have just finished sample items for the North Carolina Language Arts and Reading Assessment. Thank you for your hard work."**

STOP



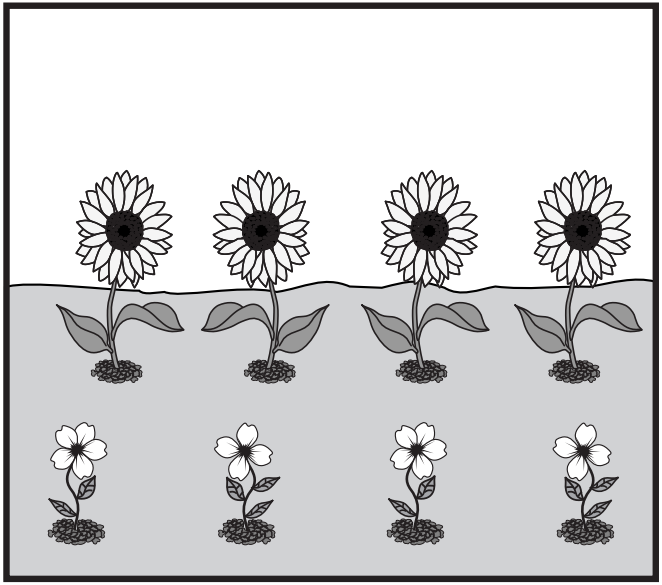
They plant seeds.

They water the seeds.

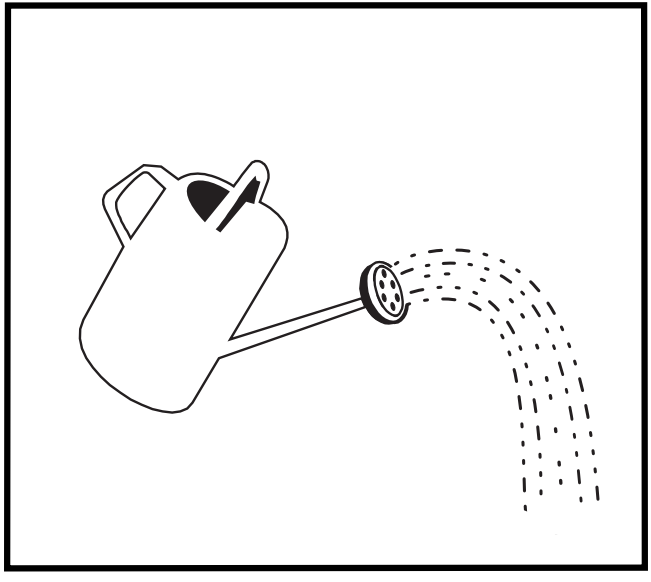


They wait for the seeds to grow.

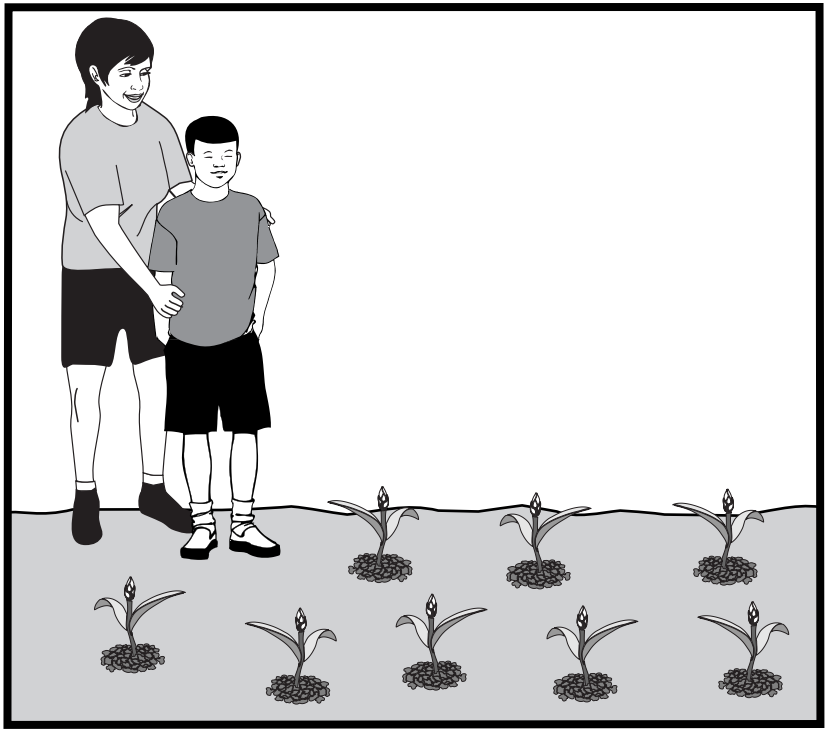
The Seeds



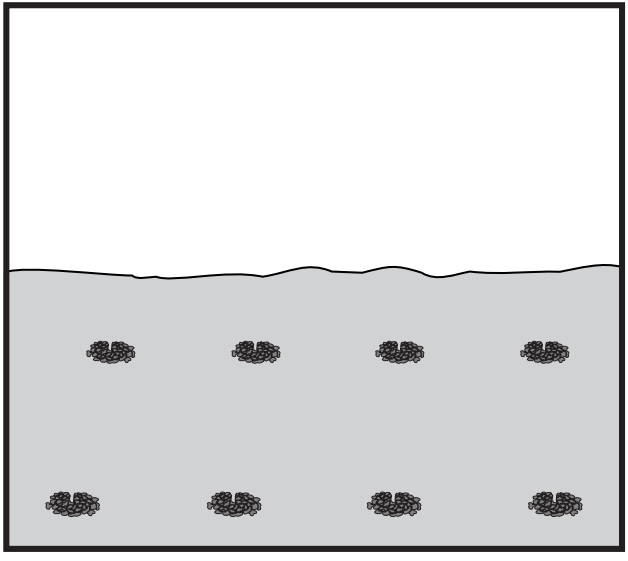
The Flowers



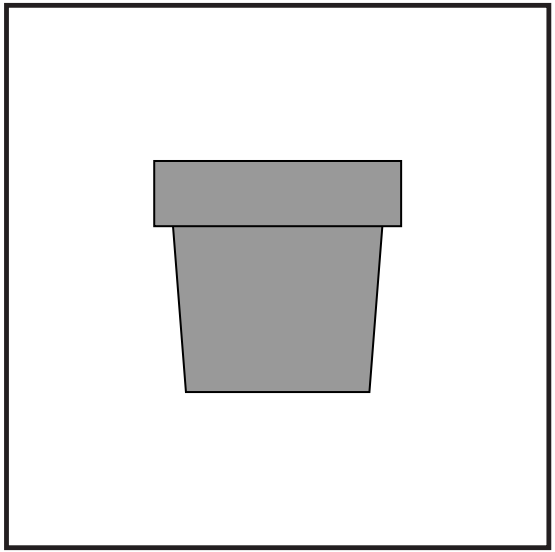
The Watering Can



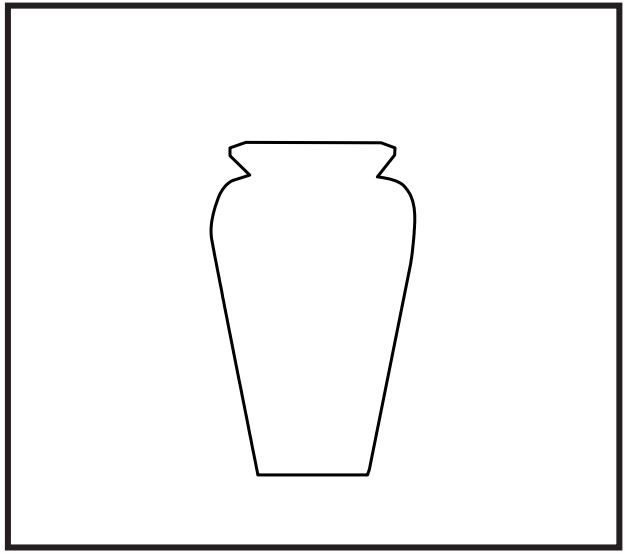
Rafael and his mother put some flowers in a _____.



garden



pot



vase

Released Form

NCEXTEND1
Mathematics Assessment—Grades 3–5

Assessor Booklet

Academic Services and Instructional Support
Division of Accountability Services

READY

Item 1

Manipulatives: Provided by NCDPI

- Stimulus: 4 – 1 = ___
- Card A: 1
- Card B: 3
- Card C: 5

*Objects/symbols may be substituted for the pictures if used routinely in the classroom. (Provided by the assessor)

Trial 1

- Present the stimulus card using the following script.
- SAY: **"What does 4 minus 1 equal?"**
- Present the response cards in the following order using the script below (*Card A*, *Card B*, *Card C*).
- SAY: (Card A) **"1"** (Card B) **"3"** (Card C) **"5"**
- SAY: **"What does 4 minus 1 equal? Choose a card."**
- Record the student response below and follow the directions to continue with Trial 2 or the next item.

Fill in Trial 1 Student Response Here:		
Card A	⬭ →	Remove Card A, Trial 2
Card B	⬭ →	Go to the next item.
Card C	⬭ →	Remove Card C, Trial 2
No Response	⬭ →	Remove Card A, Trial 2

Trial 2

- Present the stimulus card using the following script.
- SAY: **"What does 4 minus 1 equal?"**
- Present the response cards using the following script.
  If Card A was removed
- SAY: (Card B) **"3"** (Card C) **"5"**
  If Card C was removed
- SAY: (Card A) **"1"** (Card B) **"3"**
- SAY: **"What does 4 minus 1 equal? Choose a card."**
- Record the student response below and continue to the next item.

Fill in Trial 2 Student Response Here:	
Card A	⬭ → Go to the next item.
Card B	⬭ → Go to the next item.
Card C	⬭ → Go to the next item.
No Response	⬭ → Go to the next item.

Item 2

Manipulatives: Provided by NCDPI

- Card A: measuring cup $\frac{1}{2}$ full
- Card B: measuring cup $\frac{3}{4}$ full
- Card C: one measuring cup full

*Objects/symbols may be substituted for the pictures if used routinely in the classroom. (Provided by the assessor)

Trial 1

- SAY: **"Which card shows a measuring cup half full?"**
- Present the response cards in the following order (*Card A*, *Card B*, *Card C*).
- SAY: **"Which card shows a measuring cup half full? Choose a card."**
- Record the student response below and follow the directions to continue with Trial 2 or the next item.

Fill in Trial 1 Student Response Here:	
Card A ⬭ →	Go to the next item.
Card B ⬭ →	Remove Card B, Trial 2
Card C ⬭ →	Remove Card C, Trial 2
No Response ⬭ →	Remove Card C, Trial 2

Trial 2

- SAY: **"Which card shows a measuring cup half full?"**
- Present the response cards using the following script.
  If Card B was removed, present Card A, then Card C
  If Card C was removed, present Card A, then Card B
- SAY: **"Which card shows a measuring cup half full? Choose a card."**
- Record the student response below and continue to the next item.

Fill in Trial 2 Student Response Here:	
Card A	⬭ → Go to the next item.
Card B	⬭ → Go to the next item.
Card C	⬭ → Go to the next item.
No Response	⬭ → Go to the next item.

Item 3

Manipulatives: Provided by NCDPI

- Card A: pentagon
- Card B: square
- Card C: triangle

*Objects/symbols may be substituted for the pictures if used routinely in the classroom. (Provided by the assessor)

Trial 1

- SAY: **"Which shape has three angles?"**
- Present the response cards in the following order using the script below (*Card A*, *Card B*, *Card C*).
- SAY: (Card A) **"pentagon"** (Card B) **"square"** (Card C) **"triangle"**
- SAY: **"Which shape has three angles? Choose a card."**
- Record the student response below and follow the directions to continue with Trial 2 or the next item.

Fill in Trial 1 Student Response Here:	
Card A ⬭ →	Remove Card A, Trial 2
Card B ⬭ →	Remove Card B, Trial 2
Card C ⬭ →	Go to the next item.
No Response ⬭ →	Remove Card B, Trial 2

Trial 2

- Present the stimulus card using the following script.
- SAY: **"Which shape has three angles?"**
- Present the response cards using the following script.
  If Card A was removed
- SAY: (Card B) **"square"** (Card C) **"triangle"**
  If Card B was removed
- SAY: (Card A) **"pentagon"** (Card C) **"triangle"**
- SAY: **"Which shape has three angles? Choose a card."**
- Record the student response below and continue to the next item.

Fill in Trial 2 Student Response Here:	
Card A	⬭ → Go to the next item.
Card B	⬭ → Go to the next item.
Card C	⬭ → Go to the next item.
No Response	⬭ → Go to the next item.

Item 4

Manipulatives: Provided by NCDPI

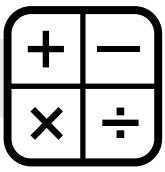
- Stimulus: analog clock showing 12:55
- Card A: 11:55
- Card B: 12:55
- Card C: 1:55

*Objects/symbols may be substituted for the pictures if used routinely in the classroom. (Provided by the assessor)

Trial 1

- Present the stimulus card using the following script.
- SAY: **"Look at this clock. What time does this clock show?"**
- Present the response cards in the following order using the script below (*Card A*, *Card B*, *Card C*).
- SAY: (Card A) **"11:55"** (Card B) **"12:55"** (Card C) **"1:55"**
- SAY: **"What time does this clock show? Choose a card."**
- Record the student response below and follow the directions to continue with Trial 2 or end the sample items.

Fill in Trial 1 Student Response Here:		
Card A	⬭ →	Remove Card A, Trial 2
Card B	⬭ →	Turn the page to end the sample items.
Card C	⬭ →	Remove Card C, Trial 2
No Response	⬭ →	Remove Card A, Trial 2



Trial 2

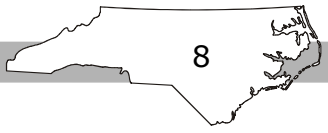
- Present the stimulus card using the following script.
- SAY: **"Look at this clock. What time does this clock show?"**
- Present the response cards using the following script.
  If Card A was removed
- SAY: (Card B) **"12:55"** (Card C) **"1:55"**
  If Card C was removed
- SAY: (Card A) **"11:55"** (Card B) **"12:55"**
- SAY: **"What time does this clock show? Choose a card."**
- Record the student response below and end the sample items.

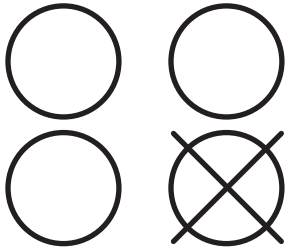
Fill in Trial 2 Student Response Here:	
Card A ⬭ →	End the sample items.
Card B ⬭ →	End the sample items.
Card C ⬭ →	End the sample items.
No Response ⬭ →	End the sample items.

End of the Sample Items

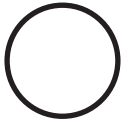
Read the following to announce the end of the testing session.

SAY: **"You have just finished sample items for the North Carolina Mathematics assessment for Grades 3–5. Thank you for your hard work."**

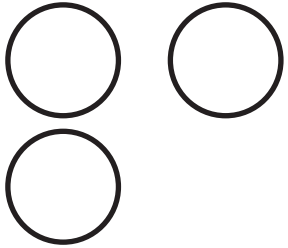




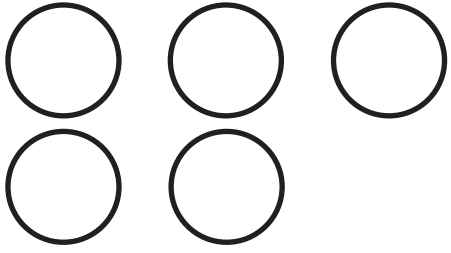
4 − 1 = _____



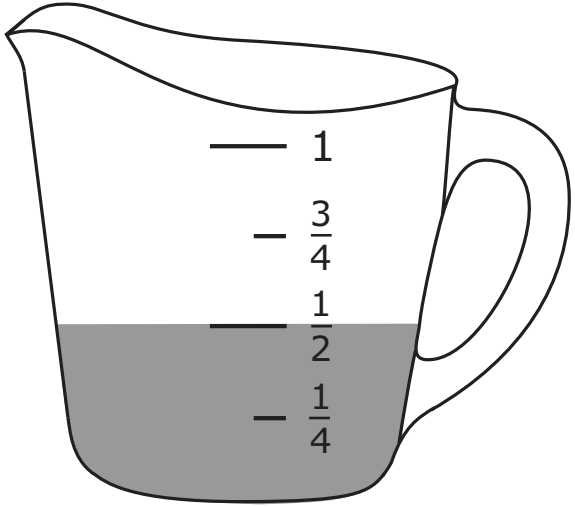
1

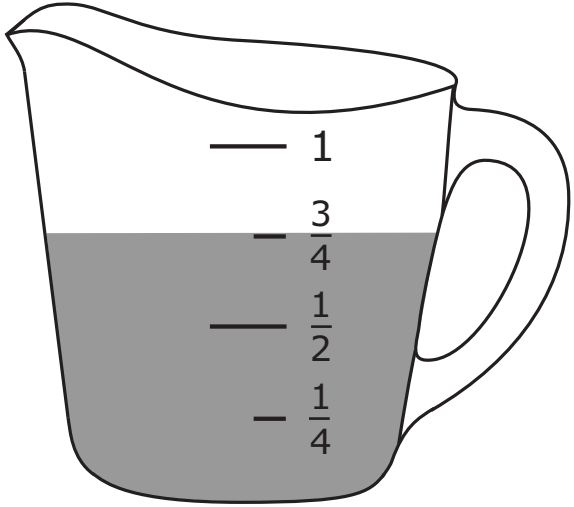


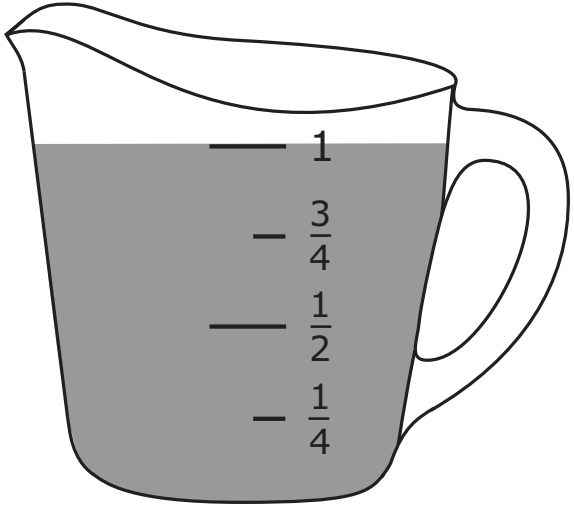
3

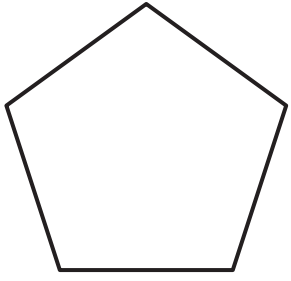


5

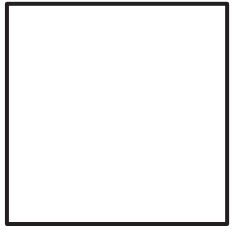

1
3/4
1/2
1/4


1
3/4
1/2
1/4

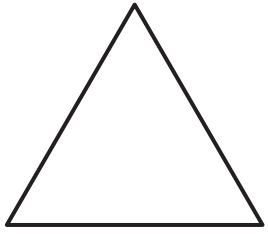

1
3/4
1/2
1/4



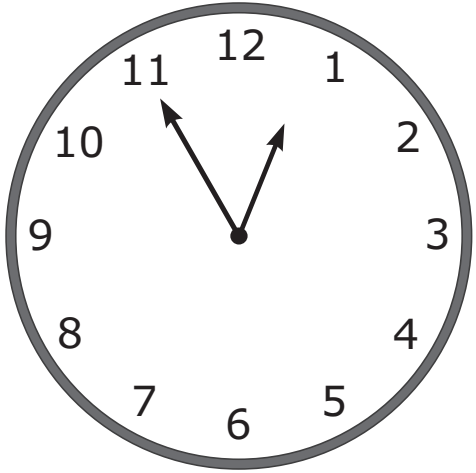
pentagon



square



triangle


12
1
2
3
4
5
6
7
8
9
10
11

11:55

12:55

1:55

Released Form

NCEXTEND1 Science Assessment—Grade 5

Assessor Booklet

Academic Services and Instructional Support
Division of Accountability Services

READY

Item 1

Manipulatives: Provided by NCDPI

- Card A: empty wheelbarrow
- Card B: wheelbarrow half full
- Card C: full wheelbarrow

*Objects/symbols may be substituted for the pictures if used routinely in the classroom. (Provided by the assessor)

Trial 1

- SAY: **"Which of these wheelbarrows is the easiest to push?"**
- Present the response cards in the following order using the script below (*Card A*, *Card B*, *Card C*).
- SAY: (Card A) **"an empty wheelbarrow"** (Card B) **"a wheelbarrow half full"** (Card C) **"a full wheelbarrow"**
- SAY: **"Which of these wheelbarrows is the easiest to push? Choose a card."**
- Record the student response below and follow the directions to continue with Trial 2 or the next item.

Fill in Trial 1 Student Response Here:		
Card A	⬭ →	Go to the next item.
Card B	⬭ →	Remove Card B, Trial 2
Card C	⬭ →	Remove Card C, Trial 2
No Response	⬭ →	Remove Card B, Trial 2

Trial 2

- SAY: **"Which of these wheelbarrows is the easiest to push?"**
- Present the response cards using the following script.
  If Card B was removed
- SAY: (Card A) **"an empty wheelbarrow"** (Card C) **"a full wheelbarrow"**
  If Card C was removed
- SAY: (Card A) **"an empty wheelbarrow"** (Card B) **"a wheelbarrow half full"**
- SAY: **"Which of these wheelbarrows is the easiest to push? Choose a card."**
- Record the student response below and continue to the next item.

Fill in Trial 2 Student Response Here:	
Card A	⬭ → Go to the next item.
Card B	⬭ → Go to the next item.
Card C	⬭ → Go to the next item.
No Response	⬭ → Go to the next item.

Item 2

Manipulatives: Provided by NCDPI

- Card A: thundering
- Card B: snowing
- Card C: raining

*Objects/symbols may be substituted for the pictures if used routinely in the classroom. (Provided by the assessor)

Trial 1

- SAY: **"What is the weather doing if frozen water is falling from the sky?"**
- Present the response cards in the following order using the script below (*Card A*, *Card B*, *Card C*).
- SAY: (Card A) **"thundering"** (Card B) **"snowing"** (Card C) **"raining"**
- SAY: **"What is the weather doing if frozen water is falling from the sky? Choose a card."**
- Record the student response below and follow the directions to continue with Trial 2 or the next item.

Fill in Trial 1 Student Response Here:	
Card A ⬭ →	Remove Card A, Trial 2
Card B ⬭ →	Go to the next item.
Card C ⬭ →	Remove Card C, Trial 2
No Response ⬭ →	Remove Card C, Trial 2

Trial 2

- SAY: **"What is the weather doing if frozen water is falling from the sky?"**
- Present the response cards using the following script.
  If Card A was removed
- SAY: (Card B) **"snowing"** (Card C) **"raining"**
  If Card C was removed
- SAY: (Card A) **"thundering"** (Card B) **"snowing"**
- SAY: **"What is the weather doing if frozen water is falling from the sky? Choose a card."**
- Record the student response below and continue to the next item.

Fill in Trial 2 Student Response Here:	
Card A	⬭ → Go to the next item.
Card B	⬭ → Go to the next item.
Card C	⬭ → Go to the next item.
No Response	⬭ → Go to the next item.

Item 3

Manipulatives: Provided by NCDPI

- Stimulus: lake
- Card A: crab
- Card B: dolphin
- Card C: fish

*Objects/symbols may be substituted for the pictures if used routinely in the classroom. (Provided by the assessor)

Trial 1

- Present the stimulus card using the following script.
- SAY: **"This is a lake. Which animal lives in a lake?"**
- Present the response cards in the following order using the script below (*Card A*, *Card B*, *Card C*).
- SAY: (Card A) **"a crab"** (Card B) **"a dolphin"** (Card C) **"a fish"**
- SAY: **"Which animal lives in a lake? Choose a card."**
- Record the student response below and follow the directions to continue with Trial 2 or end the sample items.

Fill in Trial 1 Student Response Here:		
Card A	⬭ →	Remove Card A, Trial 2
Card B	⬭ →	Remove Card B, Trial 2
Card C	⬭ →	Turn the page to end the sample items.
No Response	⬭ →	Remove Card A, Trial 2

Trial 2

- Present the stimulus card using the following script.
- SAY: **"This is a lake. Which animal lives in a lake?"**
- Present the response cards using the following script.
  If Card A was removed
- SAY: (Card B) **"a dolphin"** (Card C) **"a fish"**
  If Card B was removed
- SAY: (Card A) **"a crab"** (Card C) **"a fish"**
- SAY: **"Which animal lives in a lake? Choose a card."**
- Record the student response below and end the sample items.

Fill in Trial 2 Student Response Here:	
Card A	⬭ → End the sample items.
Card B	⬭ → End the sample items.
Card C	⬭ → End the sample items.
No Response	⬭ → End the sample items.

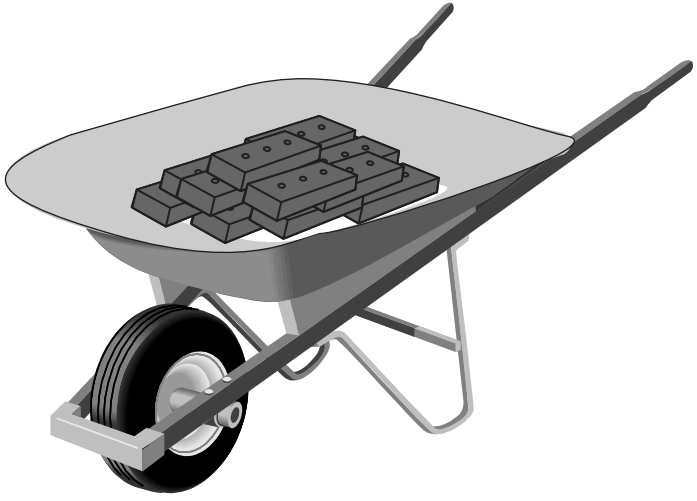
End of Sample Items

Read the following to announce the end of the testing session.

SAY: **"You have just finished sample items for the North Carolina Science assessment. Thank you for your hard work."**

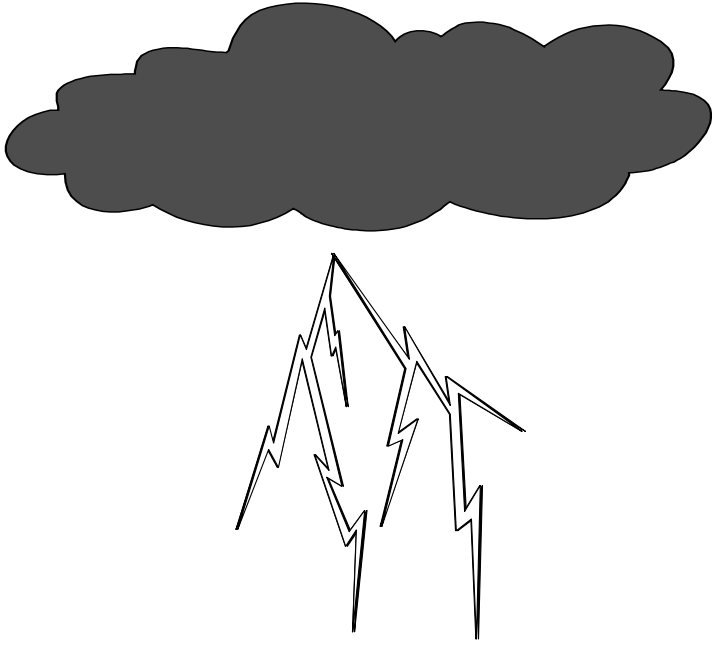
STOP

empty wheelbarrow



wheelbarrow half full

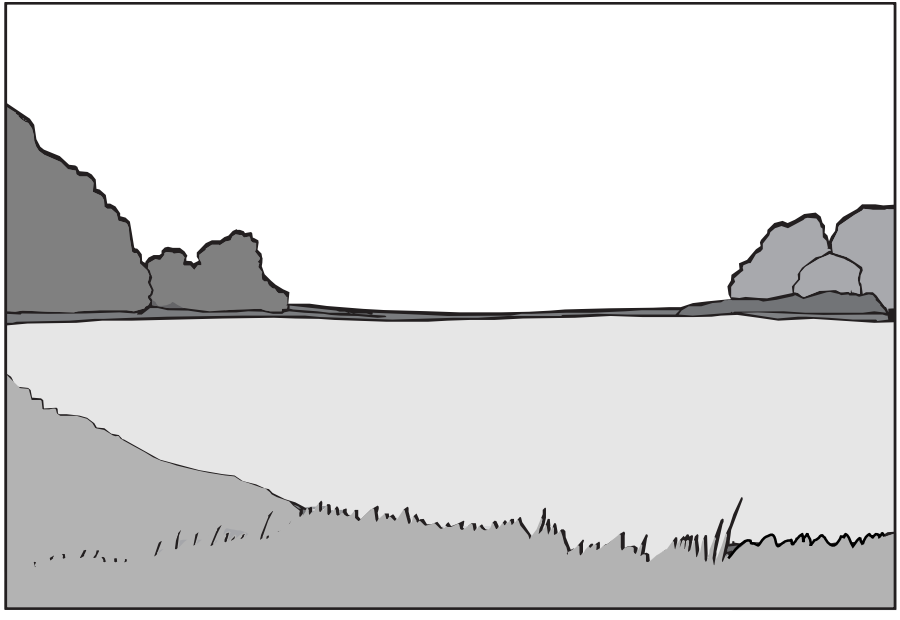
full wheelbarrow



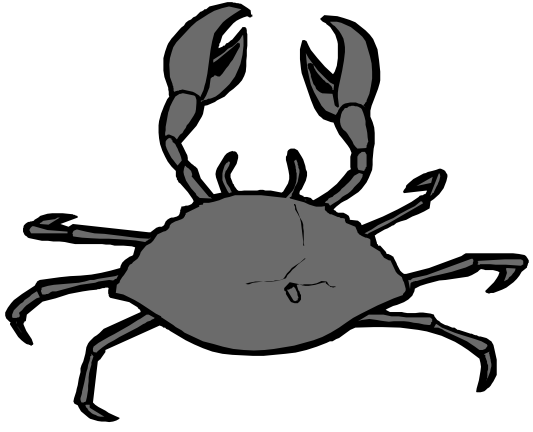
thundering

snowing

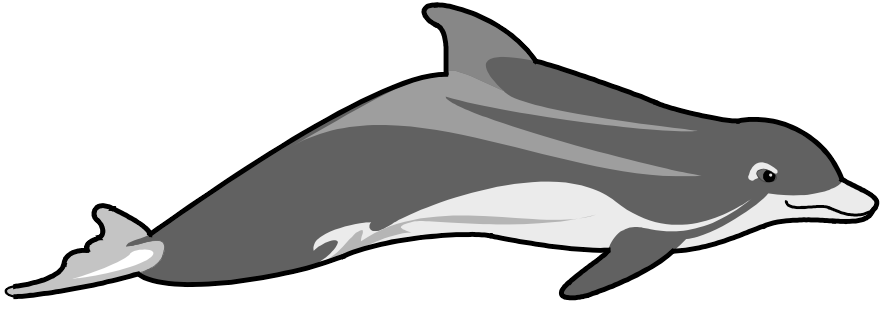
raining



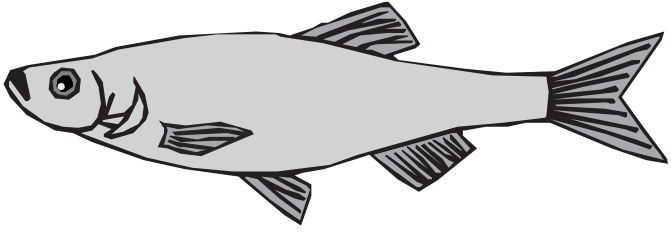
lake



crab



dolphin



fish

Appendix 4-A
*NCEXTEND*1 Form Building & Test Development Process

Item Development Process for the *NCEXTEND1* Alternate Assessment

Prior to **Step 1**, the standards to be measured must be defined. The test development process begins after new content standards are adopted by the North Carolina State Board of Education. All item writers and reviewers are required to complete training modules. The training includes a general course on item writing guidelines, including lessons on sensitivity and bias concerns. The writers and reviewers must also complete subject-specific courses on the Extend Content Standards.

Step 1: Item Created
Test items are written by North Carolina-trained item writers, including North Carolina teachers and/or curriculum specialists, and Content Specialists at Technical Outreach for Public Schools at North Carolina State University. All items are submitted on paper. The item writer assigns the item:

- an Extended Content Standard
- a secondary Extended Content Standard (when appropriate)
- a cognitive category

The item writer is also responsible for citing sources for any stimulus material to an item.

Step 2: Item Evaluation
Content Specialists review the item for accuracy of content, appropriateness of vocabulary (both subject-specific and general), adherence to item writing guidelines, and sensitivity and bias concerns. All content specialists (subject and the EC/ESL/VI specialist) look for contexts that might elicit an emotional response and inhibit students' ability to respond as well as contexts that students may be unfamiliar with for cultural or socio-economic reasons. The specialists review the item's assigned:

  - Extended Content Standard
  - secondary Extended Content Standard (if applicable)
  - Key/appropriate foils
  - difficulty rating
  - cognitive category

- If the content of the item is not accurate or does not match an objective/standard, or if the cognitive category of the item is not appropriate, the item is revised or deleted.
- If necessary, the specialist should edit the stem and foils of the items for clarity and adherence to established item writing guidelines.
- If there are necessary revisions outside the technical scope of the specialist (such as artwork, graphs, or edits to ELA selections), the item is moved to **Step 3** for edits by Production staff.
- If the item contains stimulus material, the item is moved to **Step 3** for copyright checks by Copyright staff.

Once the item is accepted, the item is sent to **Step 4** (Teacher Content Review).
The item is sent to teacher review once the content specialist has spent the needed time on the item and certifies that it is ready to be on a form.

Step 3: Production Edits/Copyright Checks
Items needing revisions outside the technical scope of the Content Specialist (such as artwork, graphs, and ELA selections) are revised by Production. Items with stimulus materials are reviewed by Copyright staff for copyright concerns and proper citation. Once the item is revised by Production or reviewed for copyrights, it is moved to **Step 2** for another review by a Content Specialist.

Step 4: Teacher Content Review
Teacher content item reviewers are required to undergo the same training as item writers. Two North Carolina-trained item reviewers look for any quality issues or bias/sensitivity issues and suggest improvements, if necessary. One of the teacher reviewers is an exceptional children's teacher, and the other is a general education teacher. The exceptional education teacher pays particular attention to the item's appropriateness for student populations with moderate to severe intellectual disabilities. Both trained reviewers evaluate the item in terms of:

- alignment to grade-level content standard
- content of item: accurate content, there is one and only one correct answer, appropriate and plausible context
- the stem is clearly written
- motivated and plausible distracters
- item design conforms to North Carolina item writing guidelines
- appropriate language for the academic content area and age of students
- bias or sensitivity concerns

Step 5: Reconcile Teacher Content Reviews
A Content Specialist carefully reviews all comments/suggestions from the content reviewers and makes any appropriate revisions. The Content Specialist may choose one of the following options:

- Send the item to **Step 6** (Production) if there are revisions required that are outside the technical scope of the Content Specialist.
- Send the item to **Step 7** (DPI-Instructional Services and EC/ESL/VI) if the item is ready for the next stage of review.
- Send it back to **Step 4** (teacher review) if major revisions are made.
- Delete the item.

Step 6: Production Edits
Items needing revisions outside the technical scope of the Content Specialist (such as artwork, graphs, and ELA selections) are revised by Production staff. Once the item is revised by Production staff, it is sent back to **Step 5** for review by a Content Specialist.

Step 7A: Curriculum-Instruction Review
The Curriculum and Instruction Specialist reviews the item and assigns an Extended Content Standard. The reviewer evaluates the item in terms of:

- alignment to grade-level extended content standard
- there is one and only one correct answer
- cognitive category
- bias, insensitivity, or accessibility issues
- overall item quality

The Curriculum and Instruction reviewer rates the item as acceptable, acceptable with revisions, or unacceptable. The review can also include additional comments. In the additional comments, the reviewer can also request that the item be returned to this step by the Test and Measurement Specialist when he or she reviews the item.

Step 7B: Exceptional Children (EC), English as a Second Language (ESL), and Visually Impaired (VI) Review
The EC/ESL/VI reviewer reviews the item for accessibility concerns for students with moderate to severe intellectual disabilities along with concerns for ESL and VI students such as accessibility of graphics for students with our without vision and also considers brailing accessibility. This review addresses concerns due to bias or insensitivity issues such as contexts that might elicit an emotional response and inhibit students' ability to respond and contexts that students may be unfamiliar with for cultural or socio-economic reasons. Review of reading level of the item is considered along with stem and foil quality (stem is a clear and complete question, foils straightforward, no repetitive words, the grammar of the stem agrees with the foils, look for idioms that may provide an accessibility issue).

Step 8: Reconcile Step 7 Reviews
A Content Specialist reviews comments/suggestions from the Curriculum and Instruction and EC/ESL/VI reviewers and makes any necessary revisions. The Specialist should indicate in the comments if any comments/suggestions from the reviewers were not approved and incorporated. The Content Specialist may choose one of the following options:

- Send the item to **Step 9** (Production) if there are revisions required that are outside the technical scope of the Content Specialist.
- Send the item to **Step 10** (TMS Review) for review.
- Send it back to **Step 4** (teacher review) if major revisions are made.
- Delete the item.

Step 9: Production Edits
Items needing revisions outside the technical scope of the Content Specialist (such as artwork, graphs, and ELA selections) are revised by Production staff. Once the item is revised by Production staff, it is sent back to **Step 8** for another review by a Content Specialist.

Step 10: Test and Measurement Specialist Review
A Test and Measurement Specialist (TMS) reviews for overall item quality. The TMS also checks that quality control measures have been followed by reading the comments from all previous reviews and verifying that the comments have been addressed by the Content Specialists.

The TMS evaluates the item for:

- alignment to grade-level content standard
- verification there is one and only one correct answer
- assigned cognitive category
- bias, insensitivity, or accessibility issues
- overall item quality

The TMS has four options when submitting the review:

- If the TMS approves the item as is, the item proceeds to **Step 13** (Grammar Review).
- If the TMS indicates edits are needed, the item proceeds to **Step 11** for review by a Content Specialist.
- If Curriculum and Instruction staff indicated they would like to see the item again, the TMS can move the item back to **Step 7** for reconciliation.
- The TMS can also choose to delete the item.

Step 11: Reconcile TMS Review, Grammar Review, or Security Review
A Content Specialist reviews comments/suggestions from the Test and Measurement Specialist from **Step 10**, Editing staff from **Step 13** (Grammar Review), or Production staff from **Step 14** (Security Review) and makes any necessary revisions. The Specialist should indicate in the comments if any comments/suggestions from the reviewers were not approved and incorporated. The Specialist may choose one of the following options:

- Send the item to **Step 12** (Production) if there are revisions required that are outside the technical scope of the Content Specialist.
- Send the item to **Step 13** (Grammar Review).
- Send it back to earlier stages of review if major revisions are made.
- Delete the item.

Step 12: Production Edits
Items needing revisions outside the technical scope of the Content Specialist (such as artwork, graphs, and ELA selections) are revised by Production staff. Once the item is revised by Production staff, it is sent back to **Step 11** for review by a Content Specialist.

Step 13: Grammar Review
The editing staff reviews the item for grammatical issues. If the item had previously been sent back to S**tep 11** by Editing, the editor should check that the suggested revisions were addressed.

- If the editor suggests revisions to the item, the item will move back to **Step 11** for review by a Content Specialist.
- If the editor approves the item as is, the item proceeds to **Step 14** (Security Check).

Step 14: Security Check
Production staff checks to make sure no duplicate copy of the item exists in previous test forms or released items. If there is a duplicate copy of the item, then the item is flagged and sent back to **Step 11**.

Step 15: Final Approval EC/ESL/VI Approval
The EC/ESL/VI specialist reviews the item to ensure it is accessible to students with moderate to severe disabilities.

Step 16: Final Approval Content Lead
The Content Lead reviews the item and makes any final necessary revisions and also reviews the item comment history to ensure all comments have been addressed. The Content Lead may choose one of the following options:

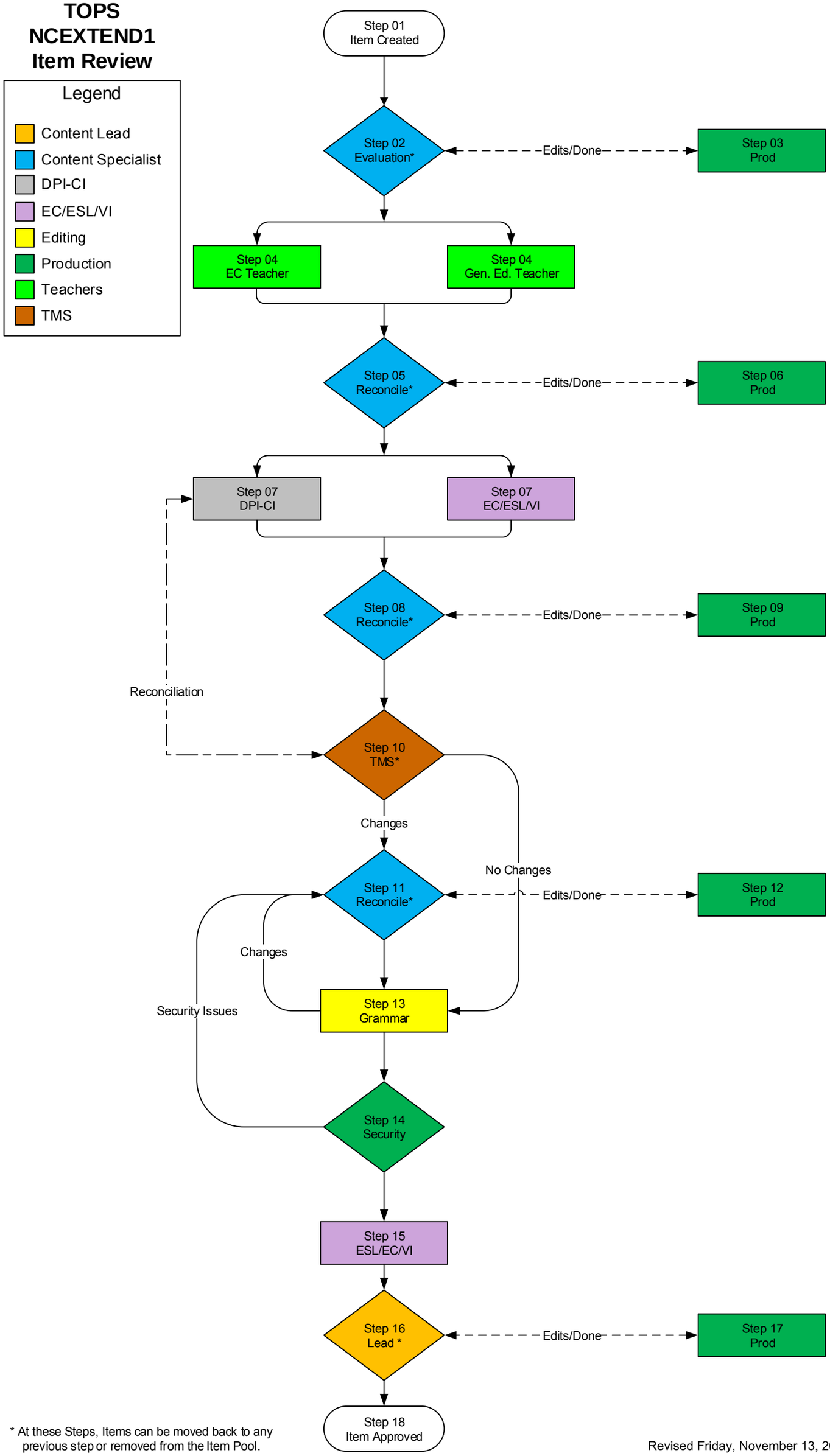
- Send the item to **Step 17** (Production) if there are revisions required that are outside the technical scope of the Content Lead.
- Approve the item and move it to **Step 18** (Item Approved).

- Send it back to **Step 2** if major revisions are made.
- Delete the item.

Step 17: Production Edits
Items needing revisions outside the technical scope of the Content Lead (such as artwork, graphs, and ELA selections) are revised by Production staff. Once the item is revised by Production staff, it is sent back to **Step 16** for review by the Content Lead.

Step 18: Item Approved
The item is now ready for placement on a form.

TOPS NCEXTEND1 Item Review



* At these Steps, Items can be moved back to any previous step or removed from the Item Pool.

Revised Friday, November 13, 2015

Selection Review Process for the *NCEXTEND1* Alternate Assessment

Prior to Step 1, an English Language Arts Content Specialist searches for appropriate selections for each assigned grade using criteria from Test Development staff, Instruction and Curriculum staff, and the *North Carolina Extended Content Standards*. The ELA Content Specialist also reviews the selections for any bias and sensitivity concerns.

Step 1: Folder Created

The Content Specialist creates a folder (color-coded by genre) for the selection. A Selection Form Submission slip is completed with the necessary copyright information (specialist's name, date, title, author, source, excerpts, etc., as well as copyright date and ISBN, if applicable) and the selection's readability score, and this is attached to the inside of the folder. Any suggested edits are noted on the selection. On the outside of the folder, a selection routing sheet is attached (includes grade level and title of selection). The Content Specialist also works with production to create graphics to illustrate content in the selections.

Step 2: Copyright Approval & Title/Author Search

The editing staff:

- Determines if the selection is public domain, gratis, or copyrighted (if copyrighted, determine whether the publisher may be used or if there is a problem, such as excessive expense).
- Searches all selection databases to determine if the selection is already in use.

Step 3: Exceptional Children (EC), English as a Second Language (ESL), and Visual Impairment (VI) Review

The EC/ESL/VI reviewer evaluates the selection for accessibility concerns for EC, ESL, and VI students in terms of:

- accessibility for students with moderate to severe intellectual disabilities
- content and length of the selection
- readability of the selection
- concerns due to bias or insensitivity issues, such as contexts that might elicit an emotional response and inhibit students' ability to respond and contexts that students may be unfamiliar with for cultural or socio-economic reasons
- accessibility of graphics for students with or without vision
- appropriateness for brailing
- prior knowledge required to understand the selection
- unfamiliar vocabulary that cannot be understood from the surrounding context

Any suggested edits are noted on the selection. Based on the review, the EC/ESL/VI reviewer can recommend to:

- use the selection
- use the selection with suggested edits
- not use the selection

Step 4: Content Lead
The Content Lead evaluates the selection in terms of:

- alignment to grade-level expectations
- content and length of the selection
- readability of the selection
- bias or sensitivity concerns
- issues brought up by copyright review

Based on review, the Content Lead can:

- approve the selection as is
- approve the selection with edits or additions (including edits to or addition of artwork); the Content Lead sends a new copy to the Copyright Staff so they can seek permission from the publisher if copyrighted
- delete the selection

Step 5: Test and Measurement Specialist Review
The Test and Measurement Specialist (TMS) evaluates the selection in terms of:

- alignment to grade-level expectations
- content and length of the selection
- readability of the selection
- bias or sensitivity concerns

The TMS also evaluates:

- any bias or sensitivity concerns raised by the EC/ESL/VI and Content Lead Reviewers
- edits made by content at **Steps 1 and 4**, or edits suggested in the **Step 4** EC/ESL/VI review

If the TMS rejects the selection, it is deleted from the pool. If the TMS approves the selection, then it moves to **Step 5**.

Step 6: Content Reconcile
Any issues noted in EC/ESL/VI and TMS reviews are reconciled by a Content Specialist.

NOTE: If any edits or additions are made to the selection (including edits to or addition of artwork), the Content Specialist sends a new copy to the Copyright Staff so they can seek permission from the publisher if copyrighted.

Step 7: Production Edits
Production staff makes edits to artwork. Once revisions are made, the selection is sent back to Step 6 for another review by a Content Specialist.

Step 8: Curriculum and Instruction Review
A Curriculum and Instruction Specialist from the Department of Public Instruction reviews the selection. The reviewer evaluates the selection in terms of:

- alignment to grade-level expectations
- content and length of the selection
- readability of the selection
- bias or sensitivity concerns

The Curriculum and Instruction Specialist rates the selection as acceptable, acceptable with revisions, or unacceptable. The Specialist can also include additional comments.

Step 9: Test and Measurement Specialist Review
The TMS does a final review on the selection and reviews all comments from the Curriculum and Instruction Specialist. The TMS either approves the selection (with comments regarding revisions, if any) or deletes the selection from the pool.

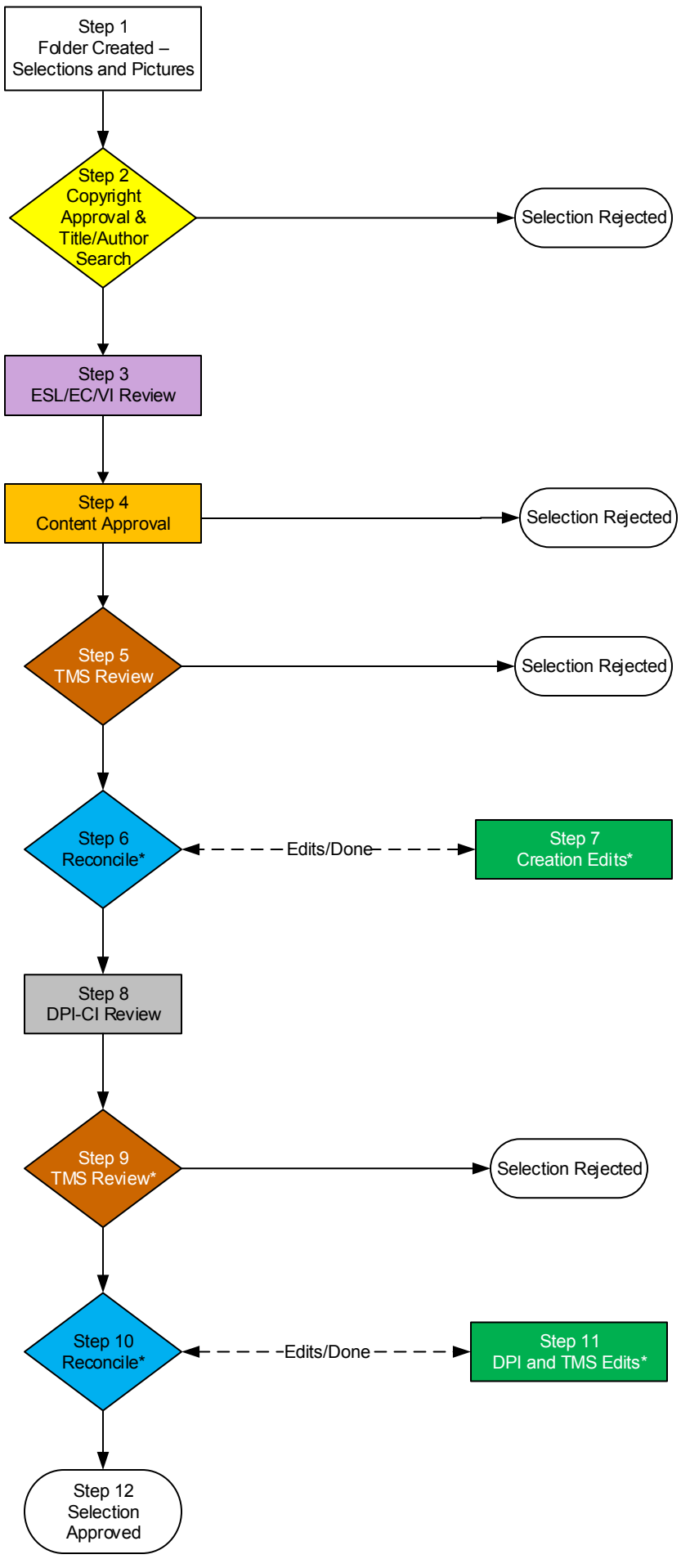
Step 10: Reconcile Curriculum and Instruction Review and Test and Measurement Specialist Review
A Content Specialist reviews any comments/changes requested by Curriculum and Instruction or by the Test and Measurement Specialist, and sends changes to **Step 7** (Production) to be made if necessary. Once any changes are made, the selection is sent to **Step 11**.

Step 12: Selection Approved
Selection is now ready to have items written.

TOPS NCEXTEND1 Selection Review

Legend

- Content Specialist
- ESL / EC / VI
- Editing
- Production
- DPI TMS
- DPI-CI
- Content Lead



* At these Steps, Selections can be moved back to any previous step or removed from the Selection Pool.

Revised Thursday, November 12, 2015 12:45:47 PM

Operational Form Review Process for the *NCEXTEND1* Alternate Assessment

Prior to Step 1: Psychometrician reviews the test items for the initial placement of the form, taking key balance into consideration.

Step 1: Select Item Numbers
A psychometrician select/approves the items to populate the form. The Psychometrician can send the form to **Step 2** (Production Edits) for revisions to artwork, graphs, or ELA selections if needed. The Psychometrician sends the form to **Step 3** for review and if needed for replacements. **Step 4 is for TMS review. TMS makes** suggestions for replacements or revisions if needed (either the content of the item or for key balancing). The Psychometrician approves any item replacement or revisions.

Step 2: Production Edits
Revisions to operational items such as artwork, graphs, and ELA selections are made by Production staff. Once the revisions are made, the form is sent back to **Step 1** for review by a Psychometrician.

Step 3: Form Review/Reconcile
A Content Specialist reviews:

- the items on the form for content alignment and quality of content, and
- the form for conflicts or repetition of content.

If any items need to be replaced due to concerns regarding conflicts or repetition of content among items, or for quality concerns, the Content Specialist sends the form back to **Step 1** with comments for the psychometrician. Otherwise, the form is sent to **Step 4** TMS review.

Step 4: Test and Measurement Specialist Review
This review step is conducted to ensure that the form is ready for Outside Content Key Check (i.e., the form is ready for students).

- This review will cover both item and form level quality.
- The Test and Measurement Specialist (TMS) will submit a review for each item, including any comments. Suggestions for revisions to items should be made only when necessary.
- After reviewing the quality of each item, the form should be evaluated in terms of cueing, repetition, and content coverage.
- The key balance of the form is checked. If the key balance is poor, the TMS will suggest which items' foils to reorder and what the key ought to be. Any suggestions for key balance edits must be approved by the Test Development Section Chief and the form is returned to **Step 1**.

After reviewing each item, the TMS can add form-level comments and suggested improvements, and can:

- send the form back to **Step 1** with suggestions for replacements or revisions,
- move the form to **Step 5** (Reconcile), or
- delete the form from the pool.

Step 5: Reconcile
At this step, the form is ready for Outside Content Key Check. The Content Specialist should review the form comments to ensure any suggested replacements or revisions have been addressed, and that any approved replacements or revisions have been made correctly. If any replacements or revisions were made incorrectly, the Content Specialist moves the form back to **Step 1** with comments. Otherwise, the form moves to **Step 7** (Outside Content Key Check).

Step 6: Production Edits
Revisions to operational items such as artwork, graphs, and ELA selections are made by Production staff. Once the revisions are made, the form is sent back to **Step 5** for review by a Psychometrician.

Step 7: Outside Content Specialist Key Check
An Outside Content Specialist reviews the form by answering each item and providing any comments and/or suggestions. This review must be done on-site.

Step 8: Reconcile Outside Content: Cueing Check and Key Balance
Content Specialist checks the keyed response from the Outside Content Review against the key for each item, and reviews all comments and/or suggestions from the Outside Content Expert. Any key disagreements are reconciled, and any comments and/or suggestions from the Outside Content Specialist are addressed.

The Content Specialist Lead, EC/ESL/VI specialist, TMS, and Content Specialist discuss comments and reviews. They check the form for cueing and ensure the key is balanced.
Not sure what else is done here.

Step 9: Reconcile
The Content ensure any suggested replacements or revisions have been addressed, and that any approved replacements or revisions have been made correctly.

Step 10: Production Edits
Revisions to items such as artwork, graphs, and ELA selections are made by Production staff. Once the revisions are made, the form is sent back to **Step 9** for review by a content specialist.

Step 11: Psychometric Review/Key Balance
A Psychometrician:

- reviews comments/suggestions from the Outside Content Specialist and from Editing staff, with consultation with the TMS and Content Specialists.
- checks key agreement with the Outside Content Specialist and resolves any disagreements through consultation with the TMS and Content Specialists.
- makes any approved revisions, or indicates revisions for Production staff to make, and sends the form to **Step 12** (Production Edits).
- checks the key balance.

Step 12: Production Edits
Revisions to items such as artwork, graphs, and ELA selections are made by Production staff. Once the revisions are made, the form is sent back to **Step 11** for review by a Psychometrician.

Step 13: Grammar Review
Two editors review the form for grammatical and/or formatting issues, providing comments and/or suggestions as needed.

Step 14: Content Specialist Review/Finalize Form
A Content Lead reviews the form and reviews all comments from Editing staff and addresses any suggestions. The Content Lead also reviews the form comment history to ensure all comments have been addressed. After reviewing the form, the Content Lead either:

- approves the form, and moves it to **Step 15 or**
- moves the form back to **Step 11** if there are edits to operational items to consider

Step 15: Production Edits
Revisions to embedded experimental items such as artwork, graphs, and ELA selections are made by Production staff. Once the revisions are made, the form is sent back to **Step 14** for review by a Content Specialist.

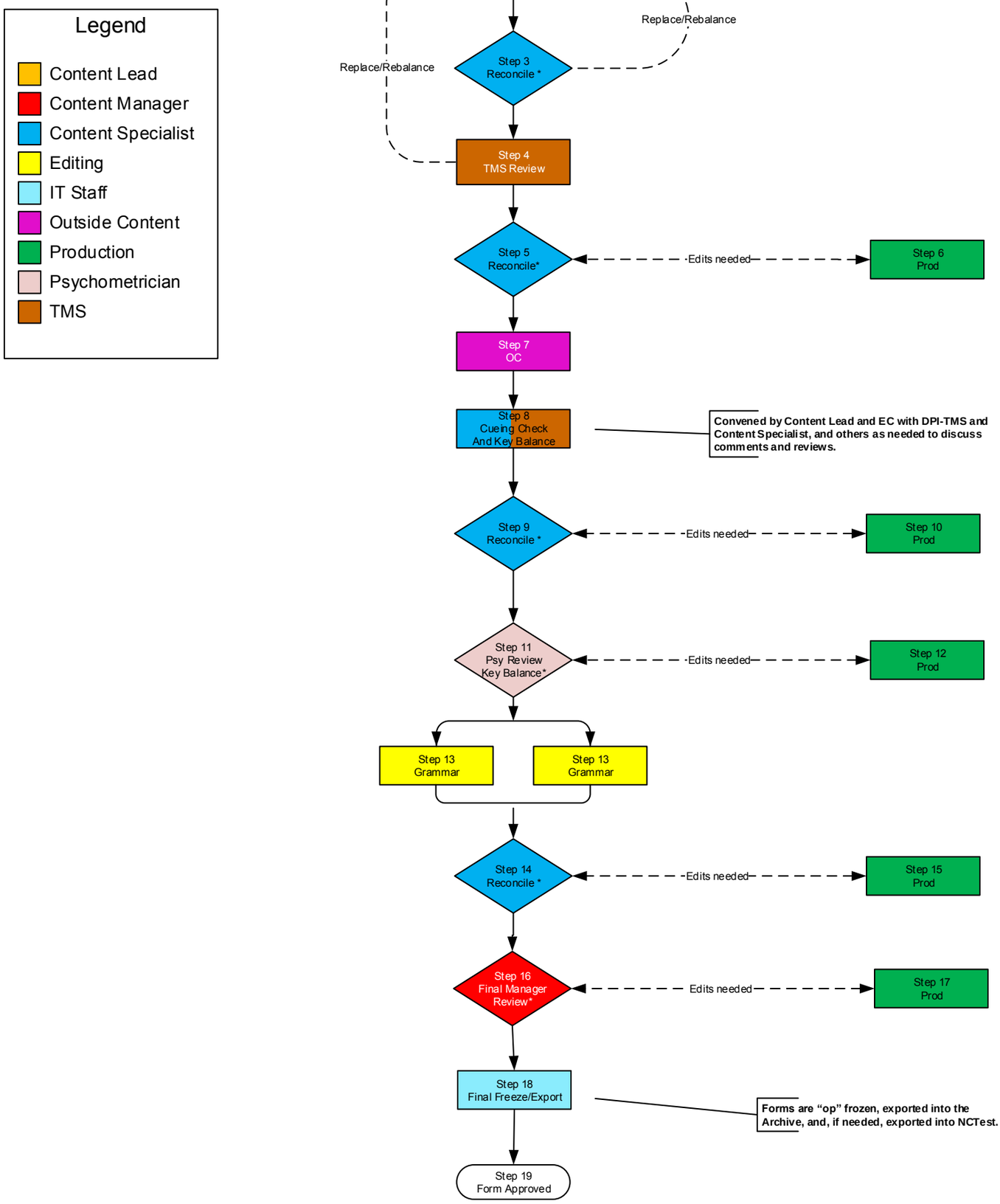
Step 16: Final Manager Review
A Content Manager reviews comments/suggestions from the Grammar Review and makes any necessary revisions to embedded items. The Manager checks the form for overall quality and reviews the form comment history to ensure all comments have been addressed.
After reviewing the form, the Content Manager may choose one of the following options:

- Approve the form and send it to **Step 18 for export to paper form**,
- Send the form to **Step 11** (Psychometrician) if there are suggested revisions to operational items for the Psychometrician to consider.
- Send the form to **Step 17** (Production Edits) for revisions to artwork, graphs, or ELA selections.
- Reject the form.

Step 17: Production Edits
Revisions to embedded experimental items such as artwork, graphs, and ELA selections are made by Production staff. Once the revisions are made, the form is sent back to **Step 16**.

Step 18: Final Export

TOPS NCEXTEND1 Form Review



*

At these Steps, Forms can be moved back to any previous step or removed from the Form Pool.

Revised Friday, November 06, 2015 11:25:42 AM

Appendix 9-A
*NCEXTEND*1 Raw Score by Subgroup

Appendix 9-A *NCEXTEND*1 Raw Score by Subgroup

Table 1. ELA Grade 3 Raw Scores by Subgroups, Population

	Grade 3 ELA	2013			2014			2015		
		N	Mean	STD	N	Mean	STD	N	Mean	STD
Ethnicity	American Indian	13	25.7	4.3	18	22.4	6.2	17	22.6	7.5
	Asian	25	20.6	7.5	30	21.5	7.2	24	20.7	6.7
	Black	332	21.5	7.9	385	22.3	6.9	406	22.3	6.4
	Hispanic	135	21.7	6.1	156	21.3	6.4	177	21.6	6.8
	Multi-Racial	49	23.3	6.3	42	21.9	7.2	52	21.8	6.5
	Native Hawiian/Pacific Islander	1	*	*	0	*	*	2	*	*
	White	434	20.9	7.4	432	22.2	7.2	542	21.8	7.1
Eco Disadvantage	Economically Disadvantaged	681	22.2	7.2	777	22.5	6.8	773	22.5	6.7
	Not Economically Disadvantaged	308	19.5	7.5	286	20.9	7.2	447	20.9	6.9
LEP	Limited English Proficient (LEP)	69	22.9	5.5	94	22.1	6.0	91	23.3	5.7
	Not LEP	920	21.3	7.5	969	22.1	7.0	1,129	21.8	6.9
Disability	Autism	367	20.6	7.0	364	22.3	6.6	436	21.9	6.4
	Deaf-Blindness	1	*	*	0	*	*	0	*	*
	Deafness	1	*	*	3	*	*	2	*	*
	Serious Emotional Disability	4	*	*	5	28.4	1.3	5	28.6	2.6
	Hearing Impairment	2	*	*	0	*	*	6	25.2	3.1
	Intellectual Disability - Mild	142	27.3	4.2	155	26.7	4.4	192	26.0	5.5
	Intellectual Disability - Moderate	209	21.9	6.1	228	22.0	5.8	267	22.0	5.5
	Intellectual Disability - Severe	50	16.3	7.3	55	16.4	5.7	44	15.5	6.9
	Specific Learning Disability	10	29.3	1.3	7	29.0	1.0	13	27.8	3.1
	Multiple Disabilities	135	16.5	8.0	142	17.4	7.5	159	16.9	7.1
	Other Health Impairment	41	23.3	7.0	68	25.3	5.9	69	24.1	6.8
	Orthopedic Impairment	2	*	*	6	27.3	4.6	4	*	*
	Speech or Language Impairment	3	*	*	2	*	*	2	*	*
	Traumatic Brain Injury	6	22.2	5.9	5	15.8	8.4	10	22.2	5.7
	Visual Impairment	2	*	*	1	*	*	3	*	*

Table 2. ELA Grade 4 Raw Scores by Subgroups, Population

	Grade 4 ELA	2013			2014			2015		
		N	**Mean**	**STD**	**N**	**Mean**	**STD**	**N**	**Mean**	**STD**
Ethnicity	American Indian	20	23.1	4.9	11	22.7	3.5	16	21.1	6.4
	Asian	26	19.1	6.3	23	19.1	6.5	30	20.0	6.0
	Black	366	20.7	5.9	329	20.7	6.6	400	20.8	6.1
	Hispanic	127	19.1	6.0	142	20.4	5.3	171	20.2	5.7
	Multi-Racial	41	20.1	6.1	49	21.2	5.3	42	21.0	5.5
	Native Hawiian/Pacific Islander	0	*	*	1	*	*	0	*	*
	White	470	20.7	6.4	463	20.4	6.2	457	20.7	6.1
Eco Disadvantage	Economically Disadvantaged	732	20.6	6.0	687	21.0	6.2	698	21.0	6.1
	Not Economically Disadvantaged	318	20.2	6.6	331	19.5	5.9	418	20.1	5.9
LEP	Limited English Proficient (LEP)	78	20.8	5.6	79	21.2	4.5	106	20.7	5.7
	Not LEP	972	20.4	6.2	939	20.5	6.3	1,010	20.7	6.1
Disability	Autism	381	19.9	6.1	373	19.5	5.9	388	20.8	5.7
	Deaf-Blindness	0	*	*	1	*	*	0	*	*
	Deafness	0	*	*	0	*	*	1	*	*
	Serious Emotional Disability	1	*	*	2	*	*	3	*	*
	Hearing Impairment	6	24.8	3.1	1	*	*	0	*	*
	Intellectual Disability - Mild	162	24.7	3.8	159	25.1	4.1	182	24.5	3.7
	Intellectual Disability - Moderate	244	20.8	4.7	235	20.8	5.1	243	20.1	5.1
	Intellectual Disability - Severe	50	16.0	5.7	48	17.9	6.0	58	14.9	5.4
	Specific Learning Disability	5	24.8	5.2	7	28.4	1.8	9	26.1	3.4
	Multiple Disabilities	138	16.6	7.2	126	17.3	7.4	140	16.6	6.2
	Other Health Impairment	43	22.9	7.0	42	23.0	5.2	75	23.7	6.3
	Orthopedic Impairment	12	24.2	4.2	1	*	*	6	23.8	6.2
	Speech or Language Impairment	0	*	*	3	*	*	1	*	*
	Traumatic Brain Injury	3	*	*	6	15.8	4.0	3	*	*
	Visual Impairment	0	*	*	2	*	*	0	*	*

Table 3. ELA Grade 5 Raw Scores by Subgroups, Population

	Grade 5 ELA	2013			2014			2015		
		N	Mean	STD	N	Mean	STD	N	Mean	STD
Ethnicity	American Indian	11	23.2	5.8	19	21.6	5.7	15	23.8	5.2
	Asian	17	20.4	7.5	27	19.9	5.5	28	20.1	5.4
	Black	363	20.3	6.4	394	20.5	5.7	367	20.3	6.5
	Hispanic	128	19.2	7.0	136	19.2	5.6	165	19.8	6.1
	Multi-Racial	44	20.2	8.0	44	20.7	5.3	47	21.8	5.2
	Native Hawiian/Pacific Islander	0	*	*	0	*	*	1	*	*
	White	516	19.9	6.5	501	20.6	6.6	508	20.1	6.5
Eco Disadvantage	Economically Disadvantaged	704	20.4	6.4	775	20.6	6.0	704	20.8	6.4
	Not Economically Disadvantaged	375	19.2	6.9	346	19.9	6.3	427	19.3	6.2
LEP	Limited English Proficient (LEP)	73	20.3	6.5	81	20.6	4.8	103	20.5	5.6
	Not LEP	1,006	20.0	6.6	1,040	20.4	6.2	1,028	20.2	6.4
Disability	Autism	353	20.1	6.4	394	20.2	5.8	410	19.4	5.9
	Deaf-Blindness	1	*	*	1	*	*	1	*	*
	Deafness	1	*	*	0	*	*	0	*	*
	Serious Emotional Disability	2	*	*	2	*	*	4	*	*
	Hearing Impairment	3	*	*	6	24.7	2.1	3	*	*
	Intellectual Disability - Mild	166	24.7	4.1	168	24.2	4.5	172	25.3	4.4
	Intellectual Disability - Moderate	267	20.6	4.6	282	20.6	5.0	267	20.0	5.0
	Intellectual Disability - Severe	65	14.0	6.2	52	14.8	6.3	57	15.2	5.7
	Specific Learning Disability	5	27.6	1.5	3	*	*	10	27.7	1.6
	Multiple Disabilities	152	15.1	7.0	142	16.8	7.0	139	16.7	7.3
	Other Health Impairment	54	22.1	7.5	39	23.6	6.1	44	24.4	4.6
	Orthopedic Impairment	2	*	*	11	22.1	5.4	3	*	*
	Speech or Language Impairment	1	*	*	1	*	*	2	*	*
	Traumatic Brain Injury	3	*	*	6	16.7	10.3	11	19.2	6.2
	Visual Impairment	1	*	*	0	*	*	2	*	*

Table 4. ELA Grade 6 Raw Scores by Subgroups, Population

	Grade 6 ELA	2013			2014			2015		
		N	Mean	STD	N	Mean	STD	N	Mean	STD
Ethnicity	American Indian	13	23.3	5.0	13	25.3	4.4	16	22.8	4.0
	Asian	22	18.8	6.8	14	22.9	4.7	34	19.6	5.4
	Black	403	20.5	6.1	374	20.3	6.0	416	20.5	6.1
	Hispanic	116	20.0	6.1	127	19.9	6.3	150	19.8	6.3
	Multi-Racial	34	20.4	4.8	43	20.0	6.9	52	20.2	4.8
	Native Hawiian/Pacific Islander	2	*	*	0	*	*	0	*	*
	White	545	20.3	5.9	532	20.4	6.1	536	20.9	6.0
Eco Disadvantage	Economically Disadvantaged	755	21.2	5.8	726	20.8	6.1	751	20.9	5.9
	Not Economically Disadvantaged	380	18.6	5.9	377	19.5	6.0	453	20.1	6.1
LEP	Limited English Proficient (LEP)	76	20.9	6.1	80	20.8	5.1	93	21.1	5.5
	Not LEP	1,059	20.3	6.0	1,023	20.3	6.2	1,111	20.6	6.0
Disability	Autism	321	19.5	5.9	362	20.1	5.8	391	20.0	6.0
	Deaf-Blindness	1	*	*	0	*	*	2	*	*
	Deafness	1	*	*	1	*	*	1	*	*
	Serious Emotional Disability	3	*	*	3	*	*	6	26.7	3.4
	Hearing Impairment	5	22.4	3.4	3	*	*	7	25.4	2.4
	Intellectual Disability - Mild	209	24.2	4.3	183	24.3	4.3	196	24.3	4.3
	Intellectual Disability - Moderate	309	20.3	4.9	278	20.6	5.2	304	20.8	4.9
	Intellectual Disability - Severe	49	13.8	6.2	57	14.3	6.3	52	15.0	5.2
	Specific Learning Disability	4	*	*	7	26.9	2.4	12	27.1	2.5
	Multiple Disabilities	153	17.6	6.5	142	16.8	6.2	157	17.1	6.4
	Other Health Impairment	53	22.8	5.9	47	22.7	6.2	49	24.1	5.1
	Orthopedic Impairment	12	20.4	5.0	6	21.7	5.9	10	23.3	3.7
	Speech or Language Impairment	0	*	*	1	*	*	0	*	*
	Traumatic Brain Injury	8	21.4	3.1	4	*	*	6	17.5	7.7
	Visual Impairment	2	*	*	1	*	*	1	*	*

Table 5. ELA Grade 7 Raw Scores by Subgroups, Population

	Grade 7 ELA	2013			2014			2015		
		N	Mean	STD	N	Mean	STD	N	Mean	STD
Ethnicity	American Indian	23	23.5	6.8	18	23.7	7.7	16	23.5	5.7
	Asian	27	19.8	5.3	26	18.5	4.8	18	20.4	4.9
	Black	379	21.0	6.2	411	20.8	6.0	393	21.0	6.2
	Hispanic	106	18.7	7.2	125	20.5	6.3	129	20.5	6.0
	Multi-Racial	24	20.1	5.6	31	22.2	5.2	49	20.4	8.1
	Native Hawiian/Pacific Islander	1	*	*	2	*	*	1	*	*
	White	510	20.3	7.0	557	20.6	6.5	543	20.9	6.3
Eco Disadvantage	Economically Disadvantaged	729	21.1	6.5	771	21.6	6.2	693	21.3	6.3
	Not Economically Disadvantaged	341	19.0	7.1	399	18.9	6.0	456	20.4	6.2
LEP	Limited English Proficient (LEP)	63	20.5	6.1	84	21.5	5.8	89	20.5	5.4
	Not LEP	1,007	20.4	6.8	1,086	20.6	6.3	1,060	21.0	6.4
Disability	Autism	284	20.1	5.9	326	19.0	5.8	358	20.6	5.7
	Deaf-Blindness	0	*	*	1	*	*	0	*	*
	Deafness	1	*	*	2	*	*	1	*	*
	Serious Emotional Disability	6	25.2	4.8	2	*	*	2	*	*
	Hearing Impairment	1	*	*	4	*	*	4	*	*
	Intellectual Disability - Mild	190	24.9	4.6	217	25.7	3.9	199	25.5	4.0
	Intellectual Disability - Moderate	325	20.6	5.5	326	20.8	5.3	315	21.1	5.5
	Intellectual Disability - Severe	52	13.5	8.0	51	14.2	5.0	59	14.9	5.8
	Specific Learning Disability	6	28.5	1.6	2	*	*	3	*	*
	Multiple Disabilities	144	15.8	7.8	151	17.4	7.1	140	16.2	6.8
	Other Health Impairment	35	23.5	6.7	56	24.6	5.0	48	23.7	6.5
	Orthopedic Impairment	6	21.8	4.6	10	19.0	6.9	6	19.8	6.5
	Speech or Language Impairment	0	*	*	1	*	*	2	*	*
	Traumatic Brain Injury	9	22.1	6.1	6	23.7	4.8	6	21.5	6.0
	Visual Impairment	3	*	*	5	24.2	4.8	0	*	*

Table 6. ELA Grade 8 Raw Scores by Subgroups, Population

	Grade 8 ELA	2013			2014			2015		
		N	Mean	STD	N	Mean	STD	N	Mean	STD
Ethnicity	American Indian	16	22.9	3.8	23	21.0	8.1	18	23.7	5.1
	Asian	26	19.2	6.0	28	19.4	3.7	37	18.2	4.8
	Black	425	20.1	5.9	435	20.0	5.5	472	20.7	5.6
	Hispanic	120	18.8	6.1	135	19.6	5.1	145	20.4	5.6
	Multi-Racial	28	19.3	6.6	29	20.4	5.3	41	20.2	5.9
	Native Hawiian/Pacific Islander	0	*	*	1	*	*	1	*	*
	White	554	19.7	5.8	584	20.3	6.0	616	20.1	5.9
Eco Disadvantage	Economically Disadvantaged	773	20.3	5.9	827	20.6	5.4	797	21.1	5.5
	Not Economically Disadvantaged	396	18.6	5.7	408	19.1	6.1	533	19.3	5.9
LEP	Limited English Proficient (LEP)	59	20.0	5.5	79	20.4	5.1	97	21.3	5.6
	Not LEP	1,110	19.8	5.9	1,156	20.1	5.8	1,233	20.3	5.8
Disability	Autism	314	19.4	5.7	321	19.6	5.3	377	19.2	5.3
	Deaf-Blindness	0	*	*	0	*	*	1	*	*
	Deafness	0	*	*	1	*	*	2	*	*
	Serious Emotional Disability	3	*	*	5	22.6	4.1	3	*	*
	Hearing Impairment	5	21.0	2.9	2	*	*	3	*	*
	Intellectual Disability - Mild	180	24.1	3.7	190	23.9	3.8	221	24.7	3.4
	Intellectual Disability - Moderate	358	19.8	4.9	408	20.6	4.6	394	20.6	5.0
	Intellectual Disability - Severe	62	14.9	5.1	69	14.4	6.5	62	16.1	4.8
	Specific Learning Disability	1	*	*	8	26.4	2.8	5	25.0	4.6
	Multiple Disabilities	178	16.3	6.9	161	16.9	6.6	179	17.6	6.8
	Other Health Impairment	45	23.7	4.2	38	23.1	5.9	53	22.5	6.7
	Orthopedic Impairment	6	20.2	5.8	7	20.6	6.5	11	18.4	8.4
	Speech or Language Impairment	0	*	*	0	*	*	1	*	*
	Traumatic Brain Injury	11	20.5	3.2	15	19.1	6.4	6	22.8	5.0
	Visual Impairment	1	*	*	2	*	*	2	*	*

Table 7. ELA Grade 10 Raw Scores by Subgroups, Population

	Grade 10 English II	2013			2014			2015		
		N	Mean	STD	N	Mean	STD	N	Mean	STD
Ethnicity	American Indian	14	18.6	3.9	19	22.6	7.9	12	19.1	10.1
	Asian	11	17.6	3.1	7	18.3	6.9	18	18.3	6.4
	Black	309	19.3	5.8	354	19.9	6.0	315	19.6	6.5
	Hispanic	77	19.3	6.0	80	18.3	6.0	94	18.6	7.5
	Multi-Racial	22	20.1	7.2	30	20.5	6.2	22	21.9	6.0
	Native Hawiian/Pacific Islander	2	*	*	0	*	*	1	*	*
	White	399	19.9	6.8	414	20.4	6.5	467	19.5	6.6
Eco Disadvantage	Economically Disadvantaged	547	20.0	6.0	606	20.5	6.4	552	19.6	7.0
	Not Economically Disadvantaged	287	18.6	6.7	298	19.2	6.0	377	19.3	6.3
LEP	Limited English Proficient (LEP)	42	19.2	6.0	38	20.1	5.2	53	19.7	5.8
	Not LEP	792	19.5	6.3	866	20.1	6.4	876	19.4	6.7
Disability	Autism	218	18.9	6.0	236	19.5	6.4	258	19.1	6.4
	Deaf-Blindness	2	*	*	1	*	*	1	*	*
	Deafness	3	*	*	2	*	*	0	*	*
	Serious Emotional Disability	3	*	*	2	*	*	4	*	*
	Hearing Impairment	1	*	*	3	*	*	3	*	*
	Intellectual Disability - Mild	112	24.2	4.6	138	24.0	4.9	145	23.9	4.9
	Intellectual Disability - Moderate	296	19.9	5.4	302	19.9	5.4	305	19.2	5.6
	Intellectual Disability - Severe	37	13.8	6.7	44	15.5	6.5	36	13.4	8.1
	Specific Learning Disability	3	*	*	0	*	*	0	*	*
	Multiple Disabilities	112	16.1	6.4	127	17.4	6.9	128	15.8	7.7
	Other Health Impairment	27	22.8	5.4	29	24.6	5.4	37	23.2	5.5
	Orthopedic Impairment	10	20.5	9.5	4	*	*	4	*	*
	Speech or Language Impairment	0	*	*	0	*	*	0	*	*
	Traumatic Brain Injury	6	16.7	6.0	8	21.4	4.5	4	*	*
	Visual Impairment	1	*	*	2	*	*	2	*	*

Table 8. Math Grade 3 Raw Scores by Subgroups, Population

	Grade 3 Math	2013			2014			2015		
		N	Mean	STD	N	Mean	STD	N	Mean	STD
Ethnicity	American Indian	13	20.2	4.7	18	19.2	6.2	17	19.7	6.5
	Asian	25	18.4	6.4	30	20.2	5.6	24	19.2	4.8
	Black	332	18.7	6.4	385	19.8	5.7	406	19.2	5.0
	Hispanic	135	19.2	4.8	156	18.6	5.3	177	18.7	5.5
	Multi-Racial	49	21.1	5.3	42	19.0	5.8	52	19.1	5.7
	Native Hawiian/Pacific Islander	1	*	*	0	*	*	2	*	*
	White	434	18.3	5.7	432	19.1	5.8	542	19.3	5.4
Eco Disadvantage	Economically Disadvantaged	681	19.3	5.8	777	19.6	5.7	773	19.5	5.2
	Not Economically Disadvantaged	308	17.5	6.0	286	18.6	5.7	447	18.7	5.4
LEP	Limited English Proficient (LEP)	69	19.3	5.2	94	19.6	5.2	91	19.6	4.4
	Not LEP	920	18.7	5.9	969	19.3	5.7	1,129	19.2	5.4
Disability	Autism	367	18.6	5.9	364	19.8	5.4	436	19.6	5.1
	Deaf-Blindness	1	*	*	0	*	*	0	*	*
	Deafness	1	*	*	3	*	*	2	*	*
	Serious Emotional Disability	4	*	*	5	24.4	3.0	5	20.8	3.5
	Hearing Impairment	2	*	*	0	*	*	6	19.5	3.4
	Intellectual Disability - Mild	142	22.1	3.9	155	21.9	4.1	192	21.9	4.4
	Intellectual Disability - Moderate	209	18.5	4.5	228	18.9	4.8	267	18.7	4.0
	Intellectual Disability - Severe	50	16.0	6.4	55	15.7	5.9	44	15.4	5.7
	Specific Learning Disability	10	25.3	3.9	7	26.9	2.4	13	23.9	3.3
	Multiple Disabilities	135	15.7	7.0	142	16.0	6.2	159	15.5	6.2
	Other Health Impairment	41	19.7	5.7	68	21.5	5.4	69	20.7	5.5
	Orthopedic Impairment	2	*	*	6	21.2	4.2	4	*	*
	Speech or Language Impairment	3	*	*	2	*	*	2	*	*
	Traumatic Brain Injury	6	18.7	3.1	5	15.4	9.9	10	20.9	3.6
	Visual Impairment	2	*	*	1	*	*	3	*	*

Table 9. Math Grade 4 Raw Scores by Subgroups, Population

	Grade 4 Math	2013			2014			2015		
		N	Mean	STD	N	Mean	STD	N	Mean	STD
Ethnicity	American Indian	19	19.8	4.1	11	19.8	3.4	16	18.1	6.8
	Asian	26	19.9	4.4	23	18.3	5.4	30	19.6	6.2
	Black	366	19.3	5.2	329	19.3	6.2	401	19.3	5.6
	Hispanic	127	17.9	5.5	142	18.9	5.3	173	19.0	5.2
	Multi-Racial	41	19.6	5.7	49	21.0	5.1	42	19.2	4.2
	Native Hawiian/Pacific Islander	0	*	*	1	*	*	0	*	*
	White	471	19.0	6.1	463	18.8	6.1	456	19.0	5.9
Eco Disadvantage	Economically Disadvantaged	732	19.1	5.5	687	19.5	6.0	699	19.4	5.6
	Not Economically Disadvantaged	318	19.1	6.0	331	18.3	5.8	419	18.6	5.6
LEP	Limited English Proficient (LEP)	78	19.1	5.2	79	19.6	4.8	106	19.3	5.5
	Not LEP	972	19.1	5.7	939	19.0	6.0	1,012	19.1	5.6
Disability	Autism	382	19.3	5.5	373	18.9	5.8	388	19.8	5.5
	Deaf-Blindness	0	*	*	1	*	*	0	*	*
	Deafness	0	*	*	0	*	*	1	*	*
	Serious Emotional Disability	1	*		2	*	*	3	*	*
	Hearing Impairment	6	20.7	2.4	1	*	*	0	*	*
	Intellectual Disability - Mild	160	22.0	4.5	159	22.9	4.2	182	21.9	4.4
	Intellectual Disability - Moderate	244	19.0	4.2	235	18.6	4.8	243	17.9	4.4
	Intellectual Disability - Severe	50	15.1	5.5	48	16.4	6.4	58	15.8	4.8
	Specific Learning Disability	5	24.6	3.0	7	25.4	4.8	9	25.8	3.2
	Multiple Disabilities	139	15.5	6.6	126	16.1	7.4	140	15.4	5.9
	Other Health Impairment	43	20.5	7.0	42	20.9	5.8	75	21.6	6.2
	Orthopedic Impairment	12	22.2	3.6	1	*	*	6	20.8	4.0
	Speech or Language Impairment	0	*	*	3	*	*	1	*	*
	Traumatic Brain Injury	3	*	*	6	16.3	5.2	3	*	*
	Visual Impairment	0	*	*	2	*	*	0	*	*

Table 10. Math Grade 5 Raw Scores by Subgroups, Population

	Grade 5 Math	2013			2014			2015		
		N	Mean	STD	N	Mean	STD	N	Mean	STD
Ethnicity	American Indian	11	22.6	4.7	19	19.7	4.3	15	21.3	3.8
	Asian	17	18.9	6.2	27	19.3	4.8	28	19.9	5.0
	Black	363	18.5	5.5	394	18.7	5.1	365	18.7	5.7
	Hispanic	128	18.0	6.0	136	18.3	4.8	164	18.7	5.0
	Multi-Racial	44	18.5	6.5	44	19.3	3.8	47	20.6	5.2
	Native Hawiian/Pacific Islander	0	*	*	0	*	*	1	*	*
	White	515	18.3	5.7	500	18.8	5.6	507	18.4	5.3
Eco Disadvantage	Economically Disadvantaged	704	18.7	5.6	775	18.7	5.1	701	18.9	5.4
	Not Economically Disadvantaged	374	17.8	5.9	345	18.8	5.4	426	18.3	5.5
LEP	Limited English Proficient (LEP)	73	18.5	5.7	81	19.6	4.2	102	19.3	4.7
	Not LEP	1,005	18.4	5.7	1,039	18.7	5.3	1,025	18.6	5.5
Disability	Autism	353	18.7	5.6	394	19.1	5.0	409	18.7	5.2
	Deaf-Blindness	1	*	*	1	*	*	1	*	*
	Deafness	1	*	*	0	*	*	0	*	*
	Serious Emotional Disability	2	*	*	2	*	*	4	*	*
	Hearing Impairment	3	*	*	6	21.3	2.9	3	*	*
	Intellectual Disability - Mild	166	21.8	4.4	168	21.0	4.3	172	21.6	4.3
	Intellectual Disability - Moderate	267	18.6	4.2	281	18.4	4.4	267	18.3	4.1
	Intellectual Disability - Severe	65	13.9	6.5	52	15.2	6.4	57	15.6	4.8
	Specific Learning Disability	5	24.4	3.0	3	*	*	9	25.0	1.6
	Multiple Disabilities	151	15.1	6.1	142	16.1	5.8	138	15.9	7.1
	Other Health Impairment	54	19.3	5.9	39	20.9	5.2	44	21.0	4.8
	Orthopedic Impairment	2	*	*	11	19.8	4.4	3	*	*
	Speech or Language Impairment	1	*	*	1	*	*	2	*	*
	Traumatic Brain Injury	3	*	*	6	17.7	9.5	11	18.0	4.9
	Visual Impairment	1	*	*	0	*	*	2	*	*

Table 11. Math Grade 6 Raw Scores by Subgroups, Population

	Grade 6 Math	2013			2014			2015		
		N	**Mean**	**STD**	**N**	**Mean**	**STD**	**N**	**Mean**	**STD**
Ethnicity	American Indian	13	21.5	4.7	13	22.8	4.5	16	20.8	6.3
	Asian	22	17.0	6.8	15	19.7	4.5	34	18.3	4.7
	Black	403	18.9	5.3	375	18.9	5.1	415	18.9	5.3
	Hispanic	116	18.7	5.9	125	19.0	5.4	149	18.4	5.3
	Multi-Racial	34	19.2	4.6	43	19.0	6.6	52	19.5	4.2
	Native Hawiian/Pacific Islander	2	*	*	0	*	*	0	*	*
	White	544	18.8	5.2	531	18.9	5.5	535	19.5	5.1
Eco Disadvantage	Economically Disadvantaged	754	19.5	5.2	726	19.3	5.3	749	19.5	5.1
	Not Economically Disadvantaged	380	17.4	5.2	376	18.4	5.5	452	18.5	5.3
LEP	Limited English Proficient (LEP)	76	19.1	5.9	79	20.0	3.8	92	19.2	4.8
	Not LEP	1,058	18.8	5.2	1,023	18.9	5.5	1,109	19.1	5.2
Disability	Autism	320	17.9	5.1	362	19.1	4.9	389	18.9	5.0
	Deaf-Blindness	1	*	*	0	*	*	2	*	*
	Deafness	1	*	*	1	*	*	1	*	*
	Serious Emotional Disability	3	*	*	3	*	*	6	25.5	2.5
	Hearing Impairment	5	21.4	1.7	3	*	*	7	21.1	3.7
	Intellectual Disability - Mild	209	22.1	3.9	184	22.0	3.8	196	21.8	4.2
	Intellectual Disability - Moderate	309	18.8	4.2	277	19.2	4.6	304	19.0	4.1
	Intellectual Disability - Severe	49	14.7	6.6	57	13.7	6.5	52	14.8	5.2
	Specific Learning Disability	4	*	*	7	23.1	1.3	12	24.5	2.6
	Multiple Disabilities	153	16.5	6.3	140	15.7	6.3	157	16.1	5.9
	Other Health Impairment	53	20.9	5.2	47	20.7	4.6	48	22.1	4.4
	Orthopedic Impairment	12	18.0	3.6	6	20.3	2.7	10	20.5	3.8
	Speech or Language Impairment	0	*	*	1	*	*	0	*	*
	Traumatic Brain Injury	8	20.1	5.1	4	*	*	6	16.8	7.4
	Visual Impairment	2	*	*	1	*	*	1	*	*

Table 12. Math Grade 7 Raw Scores by Subgroups, Population

	Grade 7 Math	2013			2014			2015		
		N	Mean	STD	N	Mean	STD	N	Mean	STD
Ethnicity	American Indian	23	19.7	7.2	18	21.7	7.3	16	24.1	5.2
	Asian	27	18.7	4.5	26	17.6	4.9	18	19.9	5.6
	Black	379	19.1	5.3	410	19.0	5.1	393	19.1	5.2
	Hispanic	105	17.9	6.3	125	19.0	5.7	129	19.2	5.0
	Multi-Racial	24	17.1	5.2	31	20.4	5.0	48	18.3	6.6
	Native Hawiian/Pacific Islander	1	*	*	2	*	*	1	*	*
	White	510	18.3	5.6	556	18.4	5.3	542	18.9	5.3
Eco Disadvantage	Economically Disadvantaged	728	19.0	5.4	769	19.6	5.3	691	19.2	5.4
	Not Economically Disadvantaged	341	17.7	5.8	399	17.2	5.1	456	18.9	5.2
LEP	Limited English Proficient (LEP)	62	18.8	5.2	84	20.0	5.5	89	19.5	4.6
	Not LEP	1,007	18.5	5.6	1,084	18.7	5.3	1,058	19.1	5.4
Disability	Autism	284	19.1	5.1	325	18.3	5.1	357	19.3	5.0
	Deaf-Blindness	0	*	*	1	*	*	0	*	*
	Deafness	1	*	*	2	*	*	1	*	*
	Serious Emotional Disability	6	20.8	5.5	2	*	*	2	*	*
	Hearing Impairment	1	*	*	4	*	*	4	*	*
	Intellectual Disability - Mild	189	21.4	4.2	217	21.7	4.2	199	21.9	3.7
	Intellectual Disability - Moderate	325	18.5	4.3	325	18.6	4.1	314	19.1	4.8
	Intellectual Disability - Severe	52	13.2	7.6	51	15.0	6.1	58	15.6	5.4
	Specific Learning Disability	6	26.5	3.4	2	*	*	3	*	*
	Multiple Disabilities	144	14.9	6.7	151	16.1	6.3	141	15.7	6.4
	Other Health Impairment	35	20.3	4.8	56	21.1	5.5	48	20.1	5.9
	Orthopedic Impairment	6	19.7	2.1	10	16.7	6.7	6	16.7	4.6
	Speech or Language Impairment	0	*	*	1	*	*	2	*	*
	Traumatic Brain Injury	9	19.7	6.4	6	20.0	9.1	6	17.5	6.6
	Visual Impairment	3	*	*	5	18.4	3.5	0	*	*

Table 13. Math Grade 8 Raw Scores by Subgroups, Population

	Grade 8 Math	2013			2014			2015		
		N	Mean	STD	N	Mean	STD	N	Mean	STD
Ethnicity	American Indian	16	19.9	4.0	23	17.5	7.0	18	20.7	4.4
	Asian	26	17.3	5.7	28	17.9	4.1	37	17.6	4.1
	Black	425	17.5	4.8	435	17.4	4.6	472	18.2	4.6
	Hispanic	120	16.4	5.1	135	17.7	4.8	145	18.2	4.8
	Multi-Racial	28	18.4	5.3	29	16.8	3.6	41	19.0	4.8
	Native Hawiian/Pacific Islander	0	*	*	1	*	*	1	*	*
	White	555	17.6	4.9	584	17.7	4.8	616	17.8	5.1
Eco Disadvantage	Economically Disadvantaged	774	17.8	5.0	827	17.9	4.5	797	18.5	4.7
	Not Economically Disadvantaged	396	16.9	4.8	408	16.8	5.1	533	17.4	5.0
LEP	Limited English Proficient (LEP)	59	17.3	4.6	79	18.2	4.9	97	18.8	4.6
	Not LEP	1,111	17.5	4.9	1,156	17.5	4.7	1,233	18.0	4.8
Disability	Autism	315	17.8	5.1	321	18.1	4.4	377	17.9	4.6
	Deaf-Blindness	0	*	*	0	*	*	1	*	*
	Deafness	0	*	*	1	*	*	2	*	*
	Serious Emotional Disability	3	*	*	5	19.6	3.8	3	*	*
	Hearing Impairment	5	20.0	3.7	2	*	*	3	*	*
	Intellectual Disability - Mild	180	19.8	3.8	190	19.7	4.0	221	21.0	3.9
	Intellectual Disability - Moderate	358	17.3	3.7	408	17.4	3.8	394	17.6	4.1
	Intellectual Disability - Severe	62	14.2	5.3	69	13.6	5.5	62	15.1	4.8
	Specific Learning Disability	1	*	*	8	24.8	4.0	5	22.4	5.0
	Multiple Disabilities	178	15.2	6.2	161	15.2	5.6	179	16.3	5.6
	Other Health Impairment	45	19.7	4.0	38	18.8	5.1	53	19.3	5.8
	Orthopedic Impairment	6	17.3	5.8	7	18.7	4.0	11	14.6	6.6
	Speech or Language Impairment	0	*	*	0	*	*	1	*	*
	Traumatic Brain Injury	11	18.0	4.0	15	17.1	5.5	6	21.0	5.9
	Visual Impairment	1	*	*	2	*	*	2	*	*

Table 14. Math Grade 10 Raw Scores by Subgroups, Population

	Grade 10 Math II	2013			2014			2015		
		N	Mean	STD	N	Mean	STD	N	Mean	STD
Ethnicity	American Indian	14	18.6	2.2	19	18.6	6.1	12	18.1	8.7
	Asian	11	16.0	2.8	7	19.1	4.3	18	17.8	6.0
	Black	309	17.7	4.8	356	18.0	4.8	315	17.5	5.7
	Hispanic	78	18.1	5.4	80	17.5	4.8	94	16.5	6.3
	Multi-Racial	22	17.9	5.4	31	18.1	5.9	22	19.1	3.9
	Native Hawiian/Pacific Islander	2	*	*	0	*	*	1	*	*
	White	399	17.4	5.6	418	18.1	5.2	467	17.9	5.2
Eco Disadvantage	Economically Disadvantaged	548	18.1	4.8	611	18.4	5.1	552	17.6	5.8
	Not Economically Disadvantaged	287	16.7	5.8	300	17.3	4.8	377	17.7	5.2
LEP	Limited English Proficient (LEP)	42	18.0	5.1	38	19.0	4.0	53	17.6	5.6
	Not LEP	793	17.6	5.2	873	18.0	5.1	876	17.7	5.6
Disability	Autism	219	17.7	5.1	239	18.1	5.3	257	18.2	5.5
	Deaf-Blindness	2	*	*	1	*	*	1	*	*
	Deafness	3	*	*	2	*	*	0	*	*
	Serious Emotional Disability	3	*	*	2	*	*	4	*	*
	Hearing Impairment	1	*	*	3	*	*	3	*	*
	Intellectual Disability - Mild	112	20.0	4.2	138	19.9	4.7	145	20.5	4.5
	Intellectual Disability - Moderate	296	17.7	4.4	305	17.6	4.1	305	17.2	4.4
	Intellectual Disability - Severe	37	13.6	6.4	45	14.8	6.0	36	13.3	7.2
	Specific Learning Disability	3	*	*	0	*	*	1	*	*
	Multiple Disabilities	112	15.6	6.2	127	16.7	5.6	128	14.5	6.5
	Other Health Impairment	27	19.3	5.1	29	21.3	4.4	37	18.9	4.7
	Orthopedic Impairment	10	18.2	6.4	4	*	*	4	*	*
	Speech or Language Impairment	0	*	*	0	*	*	0	*	*
	Traumatic Brain Injury	6	16.2	4.3	8	19.4	6.1	4	*	*
	Visual Impairment	1	*	*	2	*	*	2	*	*

Table 15. Science Grade 5 Raw Scores by Subgroups, Population

	Grade 5 Science	2013			2014			2015		
		N	Mean	STD	N	Mean	STD	N	Mean	STD
Ethnicity	American Indian	11	22.7	5.8	19	23.7	4.3	15	24.2	3.8
	Asian	17	21.4	6.2	27	20.4	3.9	28	21.0	5.0
	Black	363	20.6	6.0	391	21.3	5.2	366	20.9	6.2
	Hispanic	128	19.9	6.3	136	20.0	5.9	164	21.0	5.6
	Multi-Racial	44	20.5	7.0	44	21.3	5.3	47	22.7	4.8
	Native Hawiian/Pacific Islander	0	*	*	0	*	*	1	*	*
	White	515	20.7	6.1	497	21.2	6.1	507	20.6	6.0
Eco Disadvantage	Economically Disadvantaged	704	21.0	6.1	770	21.3	5.6	702	21.3	6.0
	Not Economically Disadvantaged	374	19.9	6.2	344	20.7	5.7	426	20.2	5.9
LEP	Limited English Proficient (LEP)	73	21.1	5.6	81	21.4	5.3	102	21.6	4.9
	Not LEP	1,005	20.6	6.2	1,033	21.1	5.7	1,026	20.8	6.0
Disability	Autism	352	20.1	5.8	391	20.8	5.4	409	20.1	5.4
	Deaf-Blindness	1	*	*	1	*	*	1	*	*
	Deafness	1	*	*	0	*	*	0	*	*
	Serious Emotional Disability	2	*	*	2	*	*	4	*	*
	Hearing Impairment	3	*	*	6	26.0	2.2	3	*	*
	Intellectual Disability - Mild	167	24.9	3.8	168	25.0	3.9	172	25.4	3.7
	Intellectual Disability - Moderate	267	21.7	4.2	280	21.5	4.6	267	21.2	4.5
	Intellectual Disability - Severe	65	15.4	6.2	52	15.8	5.6	57	16.5	6.0
	Specific Learning Disability	5	27.6	2.3	3	*	*	10	27.8	1.1
	Multiple Disabilities	152	16.3	6.8	141	17.6	6.4	138	17.1	7.7
	Other Health Impairment	53	22.3	6.6	39	22.8	6.6	44	23.6	5.4
	Orthopedic Impairment	2	*	*	11	22.8	4.3	3	*	*
	Speech or Language Impairment	1	*	*	1	*	*	2	*	*
	Traumatic Brain Injury	3	*	*	5	16.2	9.7	11	21.2	3.8
	Visual Impairment	1	*	*	0	*	*	2	*	*

Table 16. Science Grade 8 Raw Scores by Subgroups, Population

	Grade 8 Science	2013			2014			2015		
		N	Mean	STD	N	Mean	STD	N	Mean	STD
Ethnicity	American Indian	13	20.2	4.7	18	19.2	6.2	17	19.7	6.5
	Asian	25	18.4	6.4	30	20.2	5.6	24	19.2	4.8
	Black	332	18.7	6.4	385	19.8	5.7	406	19.2	5.0
	Hispanic	135	19.2	4.8	156	18.6	5.3	177	18.7	5.5
	Multi-Racial	49	21.1	5.3	42	19.0	5.8	52	19.1	5.7
	Native Hawiian/Pacific Islander	1	*	*	0	*	*	2	*	*
	White	434	18.3	5.7	432	19.1	5.8	542	19.3	5.4
Eco Disadvantage	Economically Disadvantaged	681	19.3	5.8	777	19.6	5.7	773	19.5	5.2
	Not Economically Disadvantaged	308	17.5	6.0	286	18.6	5.7	447	18.7	5.4
LEP	Limited English Proficient (LEP)	69	19.3	5.2	94	19.6	5.2	91	19.6	4.4
	Not LEP	920	18.7	5.9	969	19.3	5.7	1,129	19.2	5.4
Disability	Autism	367	18.6	5.9	364	19.8	5.4	436	19.6	5.1
	Deaf-Blindness	1	*	*	0	*	*	0	*	*
	Deafness	1	*	*	3	*	*	2	*	*
	Serious Emotional Disability	4	*	*	5	24.4	3.0	5	20.8	3.5
	Hearing Impairment	2	*	*	0	*	*	6	19.5	3.4
	Intellectual Disability - Mild	142	22.1	3.9	155	21.9	4.1	192	21.9	4.4
	Intellectual Disability - Moderate	209	18.5	4.5	228	18.9	4.8	267	18.7	4.0
	Intellectual Disability - Severe	50	16.0	6.4	55	15.7	5.9	44	15.4	5.7
	Specific Learning Disability	10	25.3	3.9	7	26.9	2.4	13	23.9	3.3
	Multiple Disabilities	135	15.7	7.0	142	16.0	6.2	159	15.5	6.2
	Other Health Impairment	41	19.7	5.7	68	21.5	5.4	69	20.7	5.5
	Orthopedic Impairment	2	*	*	6	21.2	4.2	4	*	*
	Speech or Language Impairment	3	*	*	2	*	*	2	*	*
	Traumatic Brain Injury	6	18.7	3.1	5	15.4	9.9	10	20.9	3.6
	Visual Impairment	2	*	*	1	*	*	3	*	*

Table 17. Science Grade 10 Raw Scores by Subgroups, Population

	Grade 10 Biology	2013			2014			2015		
		N	Mean	STD	N	Mean	STD	N	Mean	STD
Ethnicity	American Indian	14	20.7	4.3	19	21.0	6.9	12	19.8	9.7
	Asian	11	18.4	2.7	7	21.4	5.9	18	17.6	6.1
	Black	309	19.6	5.2	356	20.1	5.4	314	19.9	6.1
	Hispanic	78	20.1	5.6	80	18.8	5.8	94	18.8	7.1
	Multi-Racial	22	19.8	6.6	31	20.8	4.7	22	22.0	5.2
	Native Hawiian/Pacific Islander	2	*	*	0	*	*	1	*	*
	White	399	20.0	6.0	418	20.4	5.8	467	20.2	5.8
Eco Disadvantage	Economically Disadvantaged	548	20.4	5.3	611	20.6	5.8	551	20.1	6.3
	Not Economically Disadvantaged	287	18.8	6.1	300	19.4	5.2	377	19.7	5.7
LEP	Limited English Proficient (LEP)	42	20.3	5.1	38	20.7	4.9	53	19.6	6.3
	Not LEP	793	19.8	5.7	873	20.1	5.7	875	20.0	6.1
Disability	Autism	219	19.2	5.4	239	19.2	5.8	257	19.4	5.5
	Deaf-Blindness	2	*	*	1	*	*	1	*	*
	Deafness	3	*	*	2	*	*	0	*	*
	Serious Emotional Disability	3	*	*	2	*	*	4	*	*
	Hearing Impairment	1	*	*	3	*	*	3	*	*
	Intellectual Disability - Mild	112	23.2	3.9	138	23.7	4.1	145	24.1	4.1
	Intellectual Disability - Moderate	296	20.5	4.7	305	20.6	4.5	304	19.9	4.9
	Intellectual Disability - Severe	37	14.2	6.6	45	15.3	6.8	36	13.6	7.9
	Specific Learning Disability	3	*	*	0	*	*	1	*	*
	Multiple Disabilities	112	17.0	6.7	127	17.9	6.5	128	16.7	7.7
	Other Health Impairment	27	21.8	4.6	29	23.2	5.4	37	22.8	4.4
	Orthopedic Impairment	10	22.3	7.3	4	*	*	4	*	*
	Speech or Language Impairment	0	*	*	0	*	*	0	*	*
	Traumatic Brain Injury	6	20.8	2.1	8	22.0	5.0	4	*	*
	Visual Impairment	1	*	*	2	*	*	2	*	*

Appendix 10-A
Anticipated and Actual Achievement Levels

Appendix 10-A Anticipated and Actual Achievement Levels

Table 1. Agreement Count of Teacher Anticipated Achievement Level and Actual Achievement Level for NCEXTEND1 2012-2013, ELA

Grade	Actual Ach Level	ELA Anticipated Ach Level				Total
		1	2	3	4	
3	1	137	108	41	5	291
	2	61	86	39	7	193
	3	35	93	112	30	270
	4	13	58	114	32	217
4	1	111	106	52	4	273
	2	76	141	103	16	336
	3	26	53	130	21	230
	4	17	32	104	54	207
5	1	129	110	49	3	291
	2	47	84	64	4	199
	3	42	109	145	22	318
	4	14	43	139	69	265
6	1	134	121	88	7	350
	2	59	112	128	23	322
	3	17	75	180	60	332
	4	4	15	60	44	123
7	1	116	152	75	7	350
	2	34	114	144	22	314
	3	4	23	53	16	96
	4	9	54	149	101	313
8	1	130	162	95	6	393
	2	37	106	148	39	330
	3	8	65	156	60	289
	4	5	20	70	51	146
10 English II	1	130	92	42	2	266
	2	50	76	68	10	204
	3	18	54	87	26	185
	4	7	28	87	61	183

Table 2. Agreement Count of Teacher Anticipated Achievement Level and Actual Achievement Level for NCEXTEND1 2012-2013, Math

Grade	Actual Ach Level	Math Anticipated Ach Level				Total
		1	2	3	4	
3	1	127	135	38	5	305
	2	108	156	116	16	396
	3	21	63	122	25	231
	4	2	9	17	12	40
4	1	126	134	53	3	316
	2	58	153	103	11	325
	3	35	91	127	35	288
	4	4	23	58	33	118
5	1	148	164	58	6	376
	2	88	142	143	16	389
	3	17	48	105	31	201
	4	4	14	59	33	110
6	1	126	117	79	6	328
	2	92	193	196	39	520
	3	10	53	128	60	251
	4	.	2	15	13	30
7	1	128	167	111	14	420
	2	48	172	217	67	504
	3	6	16	66	37	125
	4	.	2	10	11	23
8	1	143	247	162	25	577
	2	25	134	205	52	416
	3	5	34	72	38	149
	4	2	1	6	8	17
10 Math I	1	157	117	56	5	335
	2	42	92	70	4	208
	3	24	68	106	22	220
	4	3	10	30	30	73

Table 3. Agreement Count of Teacher Anticipated Achievement Level and Actual Achievement Level for NCEXTEND1 2012-2013, Science

Grade	Actual Ach Level	Science Anticipated Ach Level				Total
		1	2	3	4	
5	1	136	100	23	3	262
	2	85	164	95	6	350
	3	25	59	98	10	192
	4	22	67	135	45	269
8	1	121	101	51	3	276
	2	56	150	137	7	350
	3	21	82	164	32	299
	4	7	55	119	53	234
10 Biology	1	108	67	33	2	210
	2	56	112	104	16	288
	3	29	54	105	17	205
	4	8	17	62	47	134

Appendix 11-A

Alpine Testing Solutions— Standard Setting Report for the North Carolina EXTEND1 Assessments

Standard Setting Report for the

North Carolina EXTEND1 Assessments

Final

August 27, 2013

Submitted by:
Chad Buckendahl
Senior Psychometrician
Director of Education Services
2467 Cordoba Bluff Ct.
Las Vegas, NV 89135
Phone/Fax: 702.586.7386
Mobile: 402.770.0085
Chad.Buckendahl@alpinetesting.com

Table of Contents

Statement of Confidentiality

The information provided in this report is proprietary and confidential. It is meant to be used by NCDPI solely for the purpose of developing the NCEXTEND1 examination. The information may not be duplicated, distributed, or used without the permission of Alpine Testing Solutions, Inc.

Acknowledgements

We would like to acknowledge several people who assisted us with the Standard Setting Workshop. Panelists of select educators make up the largest contingent of people whose work contributed to the outcome of the standard setting workshops. They participated in the activities that resulted in the cut score recommendations for each of the North Carolina EXTEND1 Assessments. The success of the workshops was due, in large part, to their efforts.

We also appreciate the assistance and support of Ms. Hope Lung, Dr. Jami-Jon Pearson, and the rest of the team from the North Carolina Department of Public Instruction during the preparation and execution of this standard setting workshop.

Purpose and Overview

The purpose of this report is to document the procedures and analyses undertaken to assist the North Carolina Department of Public Instruction (NCDPI) in recommending achievement levels descriptors and cut scores for the North Carolina EXTEND1 Assessments. The included assessments were for Reading (grades 3-8 and 10), Mathematics (grades 3-8 and 10), Science (grades 5, 8, and 10), and the Multi-Subject assessment (grade 11).

This report summarizes the procedures and the results of standard setting workshops conducted July 30-August 1, 2013. The first part of the results is the recommended Achievement Level Descriptors drafted by the standard setting panelists. These descriptors illustrate the expected knowledge, skills, and abilities of students by achievement level, grade level, and subject area. The second set of results includes the recommended cut scores for each assessment within the EXTEND1 program.

Standard Setting Report for the EXTEND1 Assessments

North Carolina EXTEND1 Assessments

The North Carolina EXTEND1 program encompasses the alternate assessments for students with the most severe cognitive and physical disabilities. The grade level curriculum and test content are built to represent the progression and continual development of knowledge and skills across the successive grade levels. Each EXTEND1 assessment in English Language Arts (ELA) and Mathematics (Math) is aligned with the NCDPI Extended Content Standards based on the *Common Core State Standards* (CCSS). Science assessments are based on the Extended Content Standards of the *North Carolina Science Essential Standards*. The results of the EXTEND1 assessments are used to evaluate students' abilities and classify them into one of four achievement levels (i.e., Levels 1, 2, 3, and 4).

Standard Setting Workshop

The standard setting workshop for the North Carolina EXTEND1 assessments was conducted July 30 – August 1, 2013 in Raleigh, NC. There were two goals of this workshop. The first goal was to produce a set of recommended achievement level descriptors that summarized the expected knowledge, skills and abilities of students at each achievement level. The second goal was to elicit recommended cut scores that define the expected performance for students within each achievement level consistent with the achievement level descriptors.

The subsequent sections of this report describe the procedures used to accomplish each of these goals. Also included in this report is a summary of the results produced from the standard setting workshops. These results should be considered as recommendations to staff members at NCDPI who will further communicate with the State Board of Education to set the final achievement level descriptors and cut scores for each achievement level across grade levels and subject areas.

Methods and Procedures

Workshop Panelists

Prior to the workshop, NCDPI provided information about eligible panelists who were then recruited by Alpine to participate in each grade span panel. Each grade span panel included 14-15 content experts from across the state (Jaeger, 1991; Raymond & Reid, 2001). Each panel represented substantial experience and included teachers who had experience with the Extended Content Standards, teachers who had experience working with students with disabilities, and general education teachers across subject areas. The experience and qualifications of the panelists are noted in Table 1 below. In addition, subsets of the elementary and middle school panelists were asked to participate in a facilitated discussion of the vertical continuity of the impact of the recommended cut scores (see description of vertical moderation process below). Specific demographic information of this subgroup is also provided in Table 1.

Table 1. Experience and qualifications of each grade-level panel

Panel	Number of Panelists	Degree			Average Years of Experience
		Bachelors	Masters	Doctorate	
Elementary	15	6	9	0	13.3
Middle School	15	6	8	1	14.3
High School	14	5	8	1	16.7
Vertical Moderation	12	4	8	0	12.75
	Experience			Gender	
	Experience with Extended Curriculum	Experience with SwD (not EC)	General Education Teacher	Male	Female
Elementary	12	1	2	3	12
Middle School	13	0	2	2	13
High School	3	8	3	3	11
Vertical Moderation	10	0	2	4	8

Workshop Orientation

On the first day of the workshop, a general orientation was held for all panelists. Hope Lung from NCDPI welcomed the group. Chad Buckendahl from Alpine Testing Solutions (Alpine) provided an orientation that covered the purpose of the workshop, the goals of the workshop, and the processes that would be used to accomplish each goal. Following the orientation, panelists worked within smaller grade span panels for the remainder of the workshop. Chad Buckendahl led the High School panel, Sarah Hughes led the Middle School panel, and Laura Brooks led the Elementary panel.

Achievement Level Descriptors

To begin creating the achievement level descriptors (ALDs), panelists were divided into table groups with representation from the diversity of the participants. Each group was assigned one or two sets of ALDs to draft based on general policy level descriptors, an example provided by NCDPI, and an example presented from another state's ALDs. In addition, the panelists were told that their ALDs should focus at the transition point or threshold from one achievement level to the next (as opposed to policy, range, or reporting ALDs). This focus was to help panelists begin to think about how students perform at the transition points between adjacent levels of achievement. Within their respective subgroup, they listed ideas for each achievement level of the types of things a student at that level could do related to the Extended Content Standards for that grade level and subject area. The draft ALDs were then transferred to an electronic format so they could be shared with each grade level panel with printed copies distributed at multiple stages of drafting. Within each panel, the ALDs were reviewed for clarity and continuity across grade levels and subject areas. As part of the ALD development process, a vertical articulation process was also included. Specifically, this included members of the elementary school grade span panel meeting with members of the middle school grade span panel to discuss the transition from grades 5 to 6 for ELA and Mathematics. Similar discussions were held with the middle school grade span

panel and the high school grade span panel for ELA and Mathematics to ensure continuity and increasing expectations across the grade levels. Feedback from these cross-panel discussions were then shared with the original grade span groups to inform any additional revisions to the ALDs. Of note with the ALDs created for the high school panel, the ELA and Mathematics ALDs were identical for Grades 10 and 11. The expectations associated with the ELA and Mathematics for Grade 11 were judged by the panel to be undifferentiated from the what was expected in the primary high school assessments in these subject areas. However, for Science, there was a shift from Biology-specific expectations in Grade 10 to Life Science-specific expectations as part of the mixed subject assessment in Grade 11.

Standard Setting

The recommended range of cut scores is based on the Extended Angoff method (Plake & Hambleton, 2001). In this process, panelists are presented with the assessment just as students would see it and are asked to make item-level judgments. For each item, they are asked to imagine the "target student" and make their best judgment as to what score the student would likely achieve on each item (0 points, 1 point, 2 points). In this application, there were three groups of target students: the student that is barely level 2, the student that is barely level 3 and barely level 4. By focusing on the transition points between the achievement levels (e.g., barely level 3 differentiates between levels 2 and 3), panelists demonstrate their expectations for students who represent the minimum level of knowledge and skills at each of the upper achievement levels. These expectations are then use to represent the minimum score required for each of the upper achievement levels (i.e., the cut scores).

Panelists recorded these judgments on specially designed rating forms which the facilitator collected and used to compute the panel-level statistics. Rating forms that included their individual recommended cut scores were returned to panelists. The facilitator also shared with the panelists the group median cut scores, the range of cut scores across the panel (including a graphical representation of the distribution), the estimated impact if the median cut scores were used (i.e., what percent of students would be classified at or above each achievement level), and the average item score from the spring 2013 administration year. In addition, the group discussed two items for each assessment – one that was generally easier for students and one that was more difficult – to help with understanding of how to apply the ALDs to the rating task. After explaining this feedback, the facilitator instructed the panelists to review their first round of ratings and make any modifications they felt necessary in their second round of ratings. The second round ratings were used to compute the final recommended cut scores.

Following ratings for all assessments, the final activity for the full group of panelists was the completion of an evaluation form designed to measure the level of confidence in the standard setting activities and their cut score recommendations. After finishing their evaluation forms, materials were collected. After the evaluations were completed, each participant was provided with a certificate of participation and the respective workshop was concluded.

Vertical Moderation Discussion

As noted above, a subset of panelists from the elementary and middle school panels then convened on the afternoon of the last day of the study to discuss the continuity across grade levels within a subject area. Chad Buckendahl facilitated this discussion which included English Language, Mathematics, and Science as separate

topics. After showing panelists the impact results from the second round of ratings, the panel discussed a number of questions regarding interpretation and explanation of the results. Some of the questions that were posed to the group during this discussion included whether the impact across grade levels for a given subject area appeared reasonable. In addition, panelists were asked whether any grade levels appeared unreasonably high or low in terms of expectations. Some of the context that was included in the discussion was the alignment of the ELA and Math assessments to the Extended Content Standards of the Common Core State Standards.

In general, panelists provided feedback suggesting that expectations from elementary to middle school and eventually high school increased at a trajectory that is steeper than the typical progression of development for students who take the EXTEND1 assessments. Further, there is a shift in cognitive complexity from more concrete to more abstract concepts in moving from elementary to middle school, particularly grades 6 to 7 in mathematics. There were some comments regarding the performance of students in the elementary grade levels in ELA being potentially higher than expected given the change in the expectations for students in the Extended Content Standards. Another point raised by panelists in the discussion was the influence of guessing on student performance. Given the design of the assessment administration, students had a reasonable probability of earning points on a given item through chance. There was consensus, almost unanimity, among the panelists that students would guess on items. This additional factor led us to consider including a guessing adjustment in the final recommendations to ensure that scores correspond with the meaning of the achievement levels.

Results

Achievement Level Descriptors

The draft achievement level descriptors are included in Appendix A by grade level and subject areas. We recommend that NCDPI evaluate these draft descriptors and make any modifications necessary for consistency. However, substantive changes to expectations would have the effect of confounding the interpretation of the cut scores because these were the ALDs that panelists used to make their recommended judgments on the assessments.

Standard setting

The standard setting included two rounds of judgments. The results for each grade level are presented in Tables 2-5 for English Language Arts, Mathematics, Science, and the Multi-Subject assessments, respectively. From the first round of ratings, each table includes the median recommended cut score (R1-Median) for each level along with the estimated impact (R1-Impact, percent of students at or above each performance level). From the second round of ratings, each table includes the median recommended cut score (R2-Median) for each level along with the estimated impact (R2-Impact, percent of students at or above each performance level), the standard deviation of the recommended cut scores (R2-SD) which represents the variability among the panel, and the range of recommended cut scores (R2-Range) which was estimated using the variability among the panel. Specifically, the range of recommended cut scores is estimated as:

High End of the Range = Median + 2 Standard Error of the Median

Low End of the Range = Median - 2 Standard Error of the Median

where,

Standard Error of the Median = 1.25 * Stdev/sqrt(N).

The full results are shown graphically in Appendix B. Specifically, these stacked dot plots display the recommended cut score for each panelist for each performance level.

Table 2. ELA Standard Setting Results by Grade and Performance Level

Level	Result	3	4	5	6	7	8	10
	R1-Median	10	8	7	11	11	12	6.5
	R1-Impact	94%	97%	96%	95%	94%	92%	97%
2	R2-Median	10	8	7	11	11	11	8.5
	R2-Impact	94%	97%	96%	95%	94%	94%	96%
	R2-SD	3.20	3.02	2.15	3.71	3.72	3.94	5.14
	R2-Range	8-12	6-10	6-8	9-13	9-13	8-14	5-12
	R1-Median	22	20	18	21	22	21	16
	R1-Impact	52%	60%	66%	52%	48%	48%	75%
3	R2-Median	20	20	17	20	22	20	18
	R2-Impact	60%	60%	72%	58%	48%	53%	61%
	R2-SD	3.22	3.58	2.6	2.47	2.31	2.23	4.45
	R2-Range	18-22	18-22	15-19	18-22	21-23	19-21	15-21
	R1-Median	27	26	26	27	25	26	23.5
	R1-Impact	34%	26%	24%	16%	34%	18%	30%
4	R2-Median	28	26	25	27	25	26	25
	R2-Impact	29%	26%	29%	16%	34%	18%	26%
	R2-SD	1.47	2.32	1.98	2.19	1.79	2.02	2.18
	R2-Range	27-29	25-27	24-26	26-28	24-26	25-27	24-26

Table 3. Mathematics Standard Setting Results by Grade and Performance Level

Level	Result	3	4	5	6	7	8	10
	R1-Median	7	7	6	8	12	12	5
	R1-Impact	96%	97%	96%	97%	93%	92%	97%
2	R2-Median	7	7	6	9	12	10	6
	R2-Impact	96%	97%	96%	96%	93%	95%	96%
	R2-SD	2.64	1.64	2.02	4.22	3.06	2.08	2.95
	R2-Range	5-9	6-8	5-7	6-12	10-14	9-11	4-8
	R1-Median	20	19	19	22	22	22	15
	R1-Impact	46%	52%	50%	32%	29%	19%	77%
3	R2-Median	20	18	19	20	23	21	15
	R2-Impact	46%	59%	50%	47%	22%	26%	77%
	R2-SD	3.87	2.72	3.17	4.22	1.77	2.2	3.3
	R2-Range	18-22	16-20	17-21	17-23	22-24	20-22	13-17
	R1-Median	26	26	25	27	27	27	22
	R1-Impact	13%	13%	15%	5%	6%	2%	21%
4	R2-Median	27	26	25	27	28	27	23
	R2-Impact	7%	13%	15%	5%	4%	2%	15%
	R2-SD	3.10	2.08	2.40	2.54	1.61	2.33	2.49
	R2-Range	25-29	25-27	23-27	25-29	27-29	25-29	21-25

Table 4. Science Standard Setting Results by Grade and Performance Level

Level	Result	5	8	10
	R1-Median	8	10	9
	R1-Impact	96%	96%	96%
2	R2-Median	9	11	9
	R2-Impact	96%	96%	96%
	R2-SD	1.87	2.90	4.53
	R2-Range	8-10	9-13	6-12
	R1-Median	21	21	19
	R1-Impact	54%	58%	62%
3	R2-Median	21	22	19
	R2-Impact	54%	54%	62%
	R2-SD	1.87	2.50	4.60
	R2-Range	20-22	20-24	16-22
	R1-Median	27	27	25
	R1-Impact	18%	28%	20%
4	R2-Median	25	27	25
	R2-Impact	30%	28%	20%
	R2-SD	1.66	1.78	2.56

	R2-Range	24-26	26-28	23-27

Table 5. Multi-Subject Standard Setting Results by Grade and Performance Level

Level	Result	11
2	R1-Median	9
	R1-Impact	96%
	R2-Median	10
	R2-Impact	95%
	R2-SD	3.00
	R2-Range	8-12
3	R1-Median	19
	R1-Impact	60%
	R2-Median	20
	R2-Impact	54%
	R2-SD	3.67
	R2-Range	18-22
4	R1-Median	26
	R1-Impact	20%
	R2-Median	26
	R2-Impact	20%
	R2-SD	1.97
	R2-Range	25-27

Standard Setting Guessing Adjustment

Given the nature of the administration and scoring of the NCEXTEND1 assessments (e.g., 3 choices to select from followed by a second chance with only two choices), there is a reasonable probability of students earning some points on this exam by simply guessing. Because the standard setting panelists were instructed to estimate how the students would perform on the items using their knowledge, skills, and abilities, without guessing, the suggested adjustment applied is based on the probability of a student earning points on those items that they would answer incorrectly due to lack of knowledge, skills, or abilities. A full description of the guessing adjustment can be found in Appendix D.

This guessing adjustment was applied consistently across grade levels and subject areas with one notable exception, Grade 10 Mathematics. In reviewing the median recommended results from the high school panel, we observed that for the Level 3 cut score (i.e., the one that communicates that students are meeting the standard), the panel's median recommendation was at the chance level – 15 points of a possible 30. After reviewing recommendations across grade levels and subject areas, we noted that this was the only recommendation that occurred at the chance or lower level. As a result, Alpine recommended a two phase guessing adjustment for the Grade 10 Mathematics Level 3 cut score. Specifically, as a first phase of the adjustment, we recommended raising the group's recommendation to chance plus one score point which resulted in a median recommendation of 16 as opposed to the group's initial recommendation of 15. Given the standard error of the median associated with the group's recommendations, this increase falls within the 95%

confidence interval for what we might expect. The second phase was to then apply the guessing adjustment described above that was applied across grade levels and subject areas. This additional step for the Grade 10 Mathematics assessment at Level 3 was intended to be consistent with expectations for meeting the standard across grade levels, but to also apply the same statistical adjustment.

The results of the guessing adjustment are shown in Tables 6-9 for English Language Arts, Mathematics, and Science, respectively. Each table shows the Round 2 recommended median recommended cut scores along with these same values adjusted for guessing and the impact (percent of students at or above a given achievement level) of both. In addition, graphical representations that show the impact of the recommended cut scores along with the adjusted cut sores are interspersed within these tables (Figures 1-4).

Table 6. ELA Recommended Cut Scores and Impact Adjusted for Guessing

Level	Result	3	4	5	6	7	8	10
2	Median	10	8	7	11	11	11	9
	Median-Adj	18	17	17	18	18	18	17
	Impact	93.80%	96.89%	95.70%	94.93%	94.22%	93.65%	95.86%
	Impact-Adj	68.70%	73.40%	72.28%	68.53%	66.70%	65.28%	68.56%
3	Median	20	20	17	20	22	20	18
	Median-Adj	23	23	20	23	24	23	21
	Impact	59.50%	59.53%	72.28%	58.04%	47.61%	53.01%	61.11%
	Impact-Adj	49.30%	41.51%	53.80%	39.95%	37.71%	37.34%	43.97%
4	Median	28	26	25	27	25	26	25
	Median-Adj	29	27	26	28	26	27	26
	Impact	29.30%	26.04%	29.37%	16.26%	33.76%	17.87%	26.24%
	Impact-Adj	21.80%	19.62%	24.43%	10.84%	28.81%	12.70%	21.87%

Figure 1. Impact of ELA Recommended Cut Scores and Guessing Adjustment (Adj)

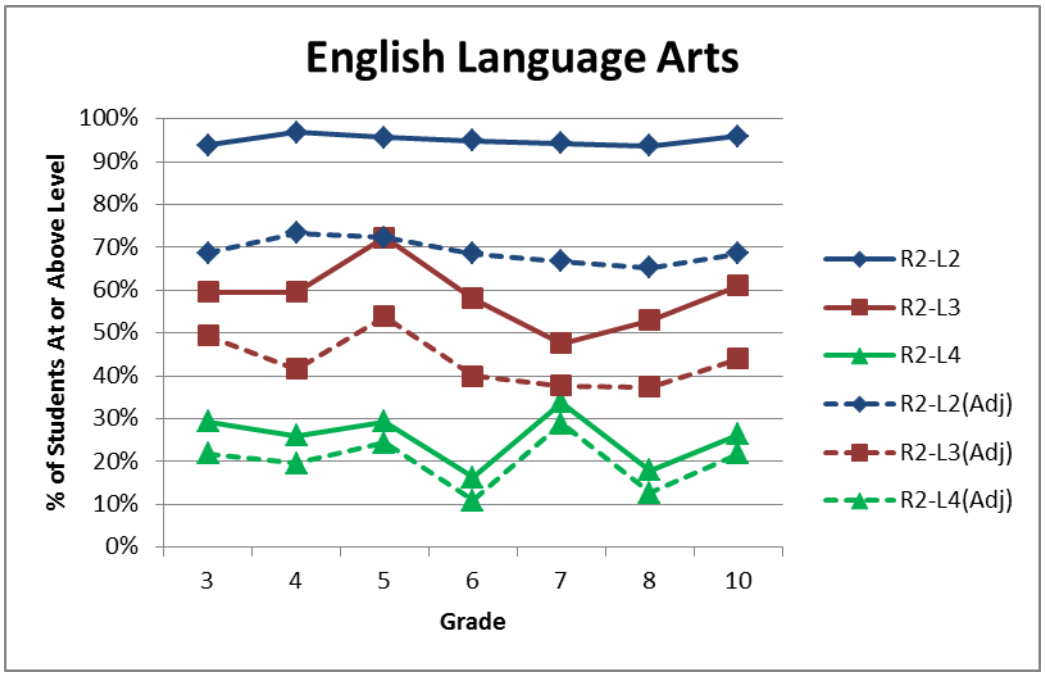


Table 7. Mathematics Recommended Cut Scores and Impact Adjusted for Guessing

Level	Result	3	4	5	6	7	8	10
2	Median	7	7	6	9	12	10	6
	Median-Adj	17	17	17	17	18	18	17
	Impact	96.00%	97.27%	96.25%	96.42%	93.03%	95.01%	96.46%
	Impact-Adj	67.60%	69.27%	64.56%	70.33%	59.95%	49.75%	59.62%
3	Median	20	18	19	20	23	21	16
	Median-Adj	23	21	22	23	25	23	20
	Impact	45.50%	58.91%	49.73%	47.03%	22.09%	25.97%	68.00%
	Impact-Adj	27.20%	38.36%	28.48%	24.52%	13.57%	14.13%	34.59%
4	Median	27	26	25	27	28	27	23
	Median-Adj	28	27	26	28	29	28	25
	Impact	7.00%	13.48%	14.84%	5.15%	3.85%	2.28%	14.99%
	Impact-Adj	4.00%	11.12%	10.07%	2.62%	2.11%	1.44%	8.62%

Figure 2. Impact of Mathematics Recommended Cut Scores and Guessing Adjustment (Adj)

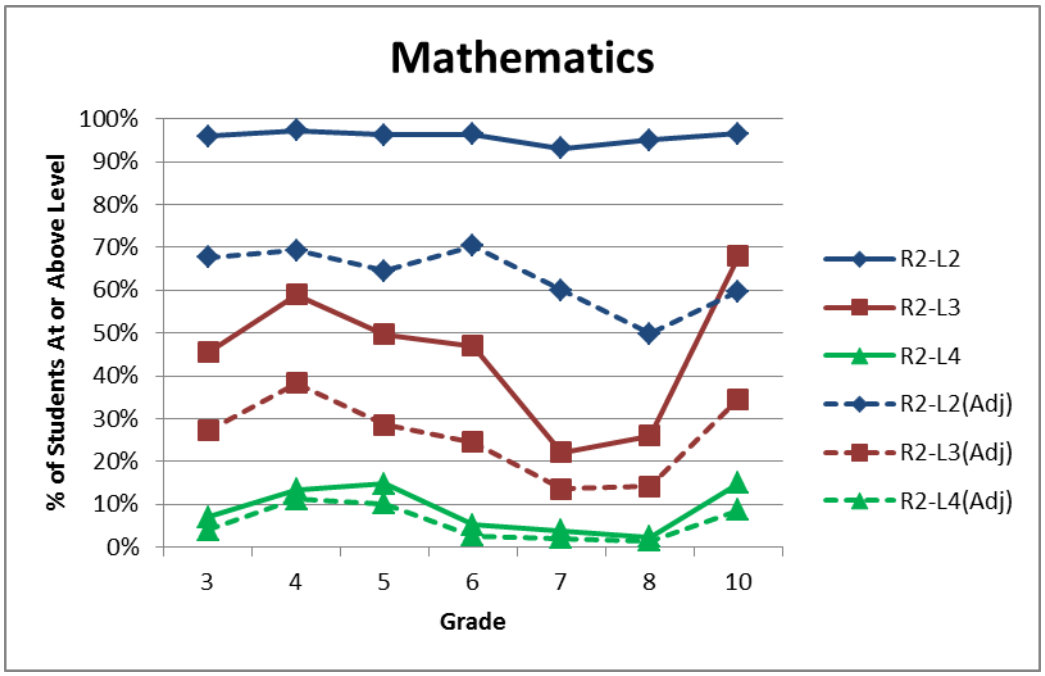


Table 8. Science Recommended Cut Scores and Impact Adjusted for Guessing

Level	Result	5	8	10
2	Median	9	11	9
	Median-Adj	17	18	17
	Impact	96.06%	95.60%	96.45%
	Impact-Adj	75.09%	75.21%	74.32%
3	Median	21	22	19
	Median-Adj	23	24	22
	Impact	54.49%	53.98%	61.66%
	Impact-Adj	42.22%	45.18%	40.24%
4	Median	25	27	25
	Median-Adj	26	28	26
	Impact	30.13%	27.66%	20.47%
	Impact-Adj	24.63%	19.80%	15.86%

Figure 3. Impact of Science Recommended Cut Scores and Guessing Adjustment (Adj)

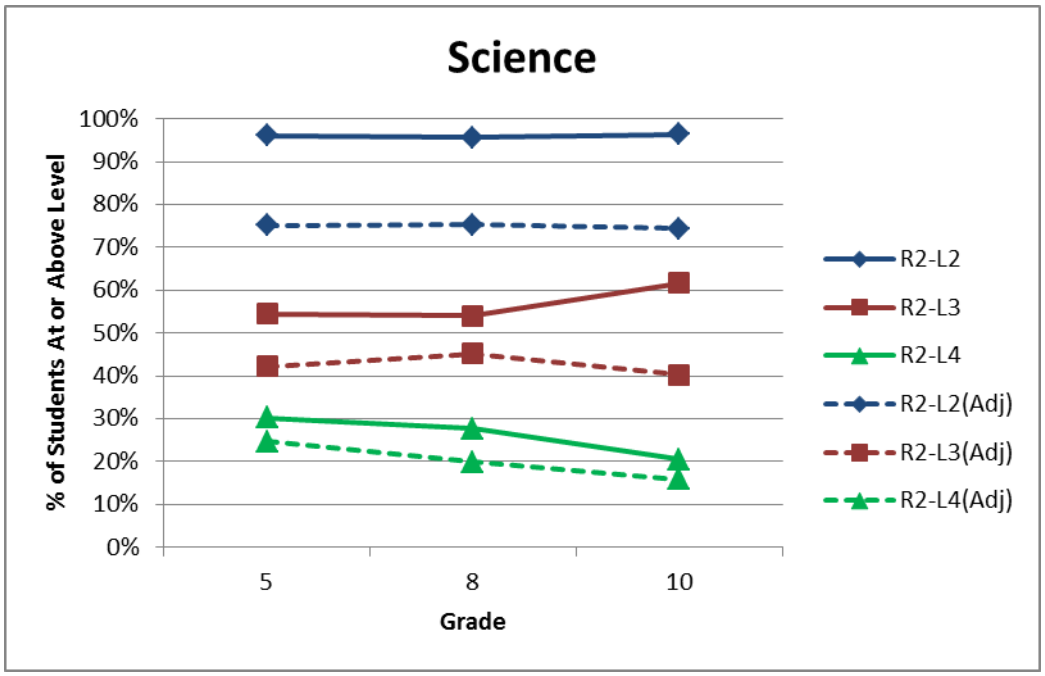
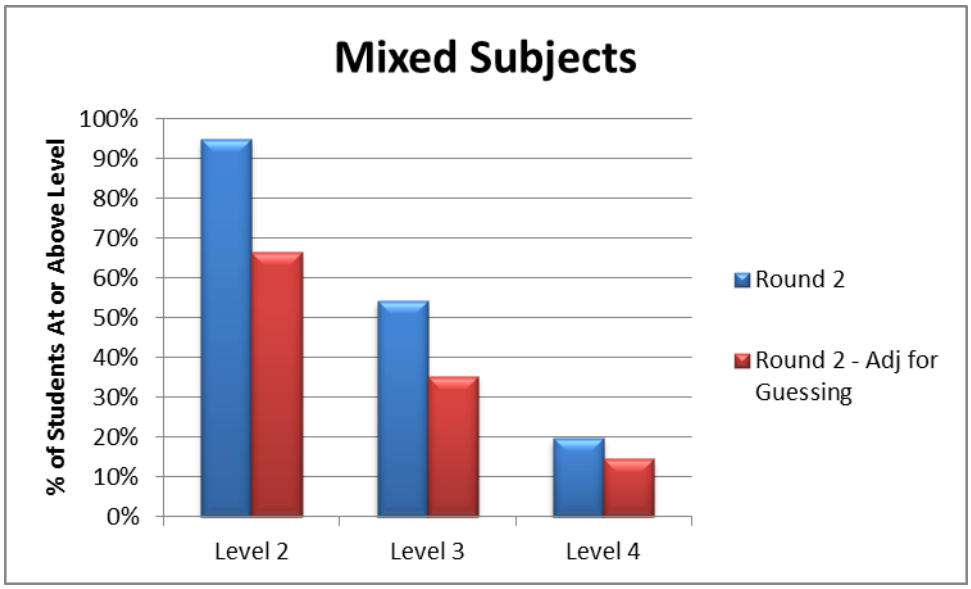


Table 9. Mixed Subjects Recommended Cut Scores and Impact Adjusted for Guessing

Level	Result	11
2	Median	10
	Median-Adj	18
	Impact	94.87%
	Impact-Adj	66.37%
3	Median	20
	Median-Adj	23
	Impact	54.17%
	Impact-Adj	35.04%
4	Median	26
	Median-Adj	27
	Impact	19.64%
	Impact-Adj	14.51%

Figure 4. Impact of Mixed Subjects Recommended Cut Scores and Guessing Adjustment (Adj)



Evaluation

Each panelist responded to a series of evaluation questions about the various components of the workshop. The median response for each panel for each evaluation question is shown in Table 10. The overall results suggest that each panel felt the workshop was very successful and felt the workshop was very successful in arriving at appropriate recommended cut scores. In addition to the closed-ended questions, panelists were allowed to provide comments about the workshop. These comments are included in Appendix E.

Table 10. Median Evaluation Responses

	Grade-Level Panel		
	Elementary	**Middle**	**High School**
Successfulness of training [6=Very Successful to 1= Very Unsuccessful]			
1a. Successfulness of orientation	5	5	5
1b. Successfulness of training on Yes/No method	5	5	4.5
1c. Successfulness of description of target students	5	5	5
1d. Successfulness of practice with method	6	5	5
1e. Successfulness of interpretation of feedback	5	5	5
1f. Successfulness of overall training	5	5	5
Time allocated to training [6= Totally Adequate to 1=Totally Inadequate]			
2a. Time – orientation	6	6	5
2b. Time – training on Yes/No method	6	4	5
2c. Time – description of target students	6	5	5
2d. Time – practice with method	6	5	5
2e. Time – interpretation of feedback	6	5	5
2f. Time – Overall training	6	5	5
Round One Yes/No Judgments			
3. Confidence in predictions [4=Confident to 1=Not at all confident]	4	3	3
4. Time for predictions [4=More than enough time to 1=More time needed]	4	3	3
Round Two Yes/No Judgments			
5. Confidence in predictions [4=Confident to 1=Not at all Confident]	4	4	4
6. Time for predictions [4=More than enough time to 1=More time needed]	4	3	3
Overall workshop			
7. Confidence in cut scores [4=Confident to 1=Not at all Confident]	3.5	4	3
8. Most useful feedback data *(mode reported)* [4=Panel summary, 3=Group discussions, 2=Impact, 1=P-values]	3	2	1
9. Least useful feedback data *(mode reported)* [4=Panel summary, 3=Group discussions, 2=Impact 1,=P-values]	4	4	3
10. Overall success [4=Very Successful to 1= Very Unsuccessful]	4	4	3

11. Overall organization [4=Very Organized to 1=Very Unorganized]	4	4	3

Conclusions

The panelists' recommendations to NCDPI and North Carolina's State Board of Education include a set of achievement level descriptors for each grade and a set of cut scores that define the performance expectations for each achievement level. We first recommend that NCDPI work with their colleagues at TOPS (NC State) to review and evaluate the achievement level descriptors after the final cut scores are set. Second, NCDPI and the State Board of Education are encouraged to consider the recommended cut scores and the positive perceptions by the panelists about their experiences and the results of the standard setting workshops.

It is important to highlight the critical elements that provide validity evidence for the results of this standard setting. Kane's (1994, 2001) framework for standard setting validity evidence identifies three elements of validity evidence for standard settings: procedural, internal, and external. Procedural validity evidence for these studies can be documented through the careful selection of representative, qualified panelists, use of a published standard setting method, completing the study in a systematic fashion, and collecting evaluation data that indicates the panelists felt they were confident in the cut score recommendations they made. Internal validity evidence suggested that panelists had similar expectations for the performance of the target students. This type of evidence is provided by the reasonable standard errors in the recommended cut scores for the second round of the standard setting process. The final type of validity evidence, external, can be provided by triangulation with results from some other estimation of appropriate cut scores from outside the current standard setting process and consideration of other factors that can influence the final policy. One way in which this could be accomplished is by conducting a second standard setting process such as contrasting groups from which one could triangulate the results of this standard setting process. From discussions with NCDPI, it appears that these data would be available to provide some additional input on the final policy decision.

References

Angoff, W. H. (1971). Scales, norms, and equivalent scores. In R. L. Thorndike (Ed.), *Educational Measurement*, (2nd ed., pp. 508-600), Washington, DC: American Council on Education.

Jaeger, R. M. (1991). Selection of judges for standard setting. *Educational Measurement: Issues and Practices*, 10(2), 3-6, 10, 14.

Kane, M. T. (1994). Validating the performance standards associated with passing scores. *Review of Educational Research*, 64, 425-461.

Kane, M. T. (2001). So much remains the same: Conception and status of validation in setting standards. In G. J. Cizek (Ed.) *Setting Performance Standards: Concepts, Methods, and Perspectives* (pp. 53-88). Mahwah, NJ: Lawrence Erlbaum Associates.

Plake, B. & Hambleton, R. (2001). The analytic judgment method for setting standards on complex performance assessments. In G. J. Cizek (Ed.*) Setting Performance Standards: Concepts, Methods, and Perspectives* (pp. 283-312). Mahwah, NJ: Lawrence Erlbaum Associates.

Raymond, M. R. & Reid, J. B. (2001). Who made thee a judge? Selecting and training participants for standard setting. In G. J. Cizek (Ed.) *Setting Performance Standards: Concepts, Methods, and Perspectives* (pp. 119-157). Mahwah, NJ: Lawrence Erlbaum & Associates, Inc.

Appendix A: Recommended Achievement Level Descriptors

English Language Arts

ELA Grade 3

Level 4: Students performing at this level have a **superior command** of the knowledge and skills contained in the Extended Content Standards of the *English Language Arts Common Core State Standards (CCSS)* in Grade 3 and are academically *well prepared to engage successfully* in further studies in this content area.

Students will consistently demonstrate the ability to:

- Given a scenario from text, interpret feelings of characters
- Identify and sequence beginning, middle and end of story presented
- Identify from which point of view the story is written
- Independently read familiar and unfamiliar words with accuracy to support comprehension
- Use a variety of sentence structures and supporting details to convey thoughts on a given topic
- Communicate for the purpose of receiving feedback to further own understanding

Level 3: Students performing at this level have a **solid command** of the knowledge and skills contained in Extended Content Standards of the *English Language Arts Common Core State Standards (CCSS)* at Grade 3 and are academically prepared to engage successfully in further studies in this content area.

Students will consistently demonstrate the ability to:

- Recall key details, characters and events in a text or selection
- Identify similarities in characters or topics among two texts
- Employ strategies to answer factual questions about a text (e.g. visual aid, revisiting text; active engagement)
- Read and comprehend simple sentences composed of CVC or common sight words
- Given a topic, compose and produce a product
- Communicate with peers and adults using multi-turn exchanges and use questioning strategies to clarify information
- Use correct plural/singular nouns, adjectives and verb tenses to achieve desired outcomes when writing or communicating
- Apply new vocabulary to make real-life connections between words and their use

Level 2: Students performing at this level have a **partial command** of the knowledge and skills contained in the Extended Content Standards of the *English Language Arts Common Core State Standards (CCSS)* at Grade 3 and *will likely need academic support to engage successfully* in further studies in this content area.

Students will consistently demonstrate the ability to:

- Given choices, make a selection to answer simple questions
- Compose sentence with subject and verb
- Communicate preferences given a series of choices
- Make simple requests to meet needs
- Identify basic sight words
- Identify words to complete a sentence

Level 1: Students performing at this level have a **limited command** of the knowledge and skills contained in the Extended Content Standards of the *English Language Arts Common Core State Standards (CCSS)* at Grade 3 and *will need academic support to engage successfully* in further studies in this content area.

Students will consistently demonstrate the ability to:

- Offer indication of attending to text
- When prompted, express thoughts using single word or object response
- Follow teacher model to make a selection
- Communicate needs/wants through nonconventional needs
- Match words to complete a sentence
- Match basic sight words

ELA Grade 4

Level 4: Students performing at this level have a **superior command** of the knowledge and skills contained in the Extended Content Standards of the *English Language Arts Common Core State Standards (CCSS)* in Grade 4 and are academically *well prepared to engage successfully* in further studies in this content area.

Students will consistently demonstrate the ability to:

- Explain how details relate to the main idea in the text
- Identify key features of a variety of text types
- Compare and contrast personal experience with one experience in the text
- Assimilate information from a variety of sources to support an opinion (text, conversations, etc.)
- Given a passage or selection, independently read with accuracy to support comprehension
- Explain an opinion to persuade an audience

Level 3: Students performing at this level have a **solid command** of the knowledge and skills contained in Extended Content Standards of the *English Language Arts Common Core State Standards (CCSS)* at Grade 4 and are academically prepared to engage successfully in further studies in this content area.

Students will consistently demonstrate the ability to:

- Locate and identify details and key information in a text that supports the main topic
- Identify a variety of text types (e.g. poem, play)
- Identify details and key information
- Relate text to personal experience
- Compare and contrast two texts on the same topic
- Use letter sound knowledge and context clues when encountering unfamiliar works in a text
- Select a topic and generate ideas and details to support their opinion
- Use correct comparative and superlative adjectives, prepositions and possessive pronouns to achieve desired outcomes when writing or communicating
- Use newly acquired vocabulary to complete sentences or in context across the content areas

Level 2: Students performing at this level have a **partial command** of the knowledge and skills contained in the Extended Content Standards of the *English Language Arts Common Core State Standards (CCSS)* at Grade 4 and *will likely need academic support to engage successfully* in further studies in this content area.

Students will consistently demonstrate the ability to:

- Identify the main topic
- Given a teacher model, will match a variety of text types
- Describe a personal experience
- Provide similarities between two texts on the same topic
- Use initial consonant sound similarities to read unknown words

- Select a topic and give opinion
- Given possessive pronouns, determine ownership
- Match new vocabulary to meaning and/or picture

Level 1: Students performing at this level have a **limited command** of the knowledge and skills contained in the Extended Content Standards of the *English Language Arts Common Core State Standards (CCSS)* at Grade 4 and *will need academic support to engage successfully* in further studies in this content area.

Students will consistently demonstrate the ability to:

- Track text with gaze or other tracking tool
- Chose preferences from a variety of text types
- Match objects to personal experience
- Identify text or other visual stimuli to a given topic
- Imitate letter-sound connections
- Choose preferences from a variety of topics
- Manipulates objects to show spatial concepts
- Recognizes that pictures are representative of tangible objects

ELA Grade 5

Level 4: Students performing at this level have a **superior command** of the knowledge and skills contained in the Extended Content Standards of the *English Language Arts Common Core State Standards (CCSS)* in Grade 5 and are academically *well prepared to engage successfully* in further studies in this content area.

Students will consistently demonstrate the ability to:

- Use multimedia elements from the text to make inferences about the problem
- Categorize information that is relevant to text-based topics
- Given a short passage or selection, independently read with accuracy to demonstrate comprehension
- Use multiple word combinations to provide facts and details to support opinion
- Write a narrative providing at least 3 sequential events and a sense of closure
- Participate in discussions with peers or adults by taking turns and then summarizing the key points of others
- Apply correct grammar, punctuation and spelling patterns when writing

Level 3: Students performing at this level have a **solid command** of the knowledge and skills contained in Extended Content Standards of the *English Language Arts Common Core State Standards (CCSS)* at Grade 5 and are academically prepared to engage successfully in further studies in this content area.

Students will consistently demonstrate the ability to:

- Given a text selection, identify the problem using specific details (i.e. quotes)
- Use correct academic/content vocabulary to communicate in speaking or writing to demonstrate understanding
- Given specific evidences from a text, determine which evidence supports which topic
- Apply word analysis skills to decode and read
- Use 2-3 word combinations to provide reasons, facts or details to support opinion
- Compose a simple narrative with at least 3 events in sequence using 2-3 word combinations
- Participate in discussions with a communication partner by listening to and sharing information
- Use correct conjunctions, plural/singular nouns with matching verbs, and correct verb tenses to achieve desired outcome when writing or communicating
- Employ newly acquired vocabulary from content areas in speaking and writing

Level 2: Students performing at this level have a **partial command** of the knowledge and skills contained in the Extended Content Standards of the *English Language Arts Common Core State Standards (CCSS)* at Grade 5 and *will likely need academic support to engage successfully* in further studies in this content area.

Students will consistently demonstrate the ability to:

- Identify the problem within a text

- Match newly acquired content specific vocabulary within a topic
- Determine one text based evidence that supports the main idea of the text
- Uses letter-sound connection to read words
- Form an opinion on a given topic and provide one supporting reason
- Complete 3 events in sequence to form a narrative
- Communicate using multiple turns with communications partner
- Utilize simple verb-noun subject- predicate patterns to convey ideas

Level 1: Students performing at this level have a **limited command** of the knowledge and skills contained in the Extended Content Standards of the *English Language Arts Common Core State Standards (CCSS)* at Grade 5 and *will need academic support to engage successfully* in further studies in this content area.

Students will consistently demonstrate the ability to:

- When provided choices, will choose the problem from the story
- Choose picture related to a topic (i.e. “show me a picture of a thunderstorm”)
- Attends to text when read to
- Indicate an opinion (i.e. preference for item, pushing something away when he/she doesn’t like it)
- Respond to a communication partner (may use eye gaze, gestures, switch, etc.)
- Match picture of a familiar settings/environment in order to print communicate a real life experience to complete a sentence (i.e. “I like to…”)
- Produce name on command on assignments

ELA Grade 6

Level 4: Students performing at this level have a **superior command** of the knowledge and skills contained in the Extended Content Standards of the *English Language Arts Common Core State Standards (CCSS)* in Grade 6 and are academically *well prepared to engage successfully* in further studies in this content area.

A student performing at Grade 6 will use adapted age appropriate materials to complete tasks including:

- Determine characters response to problems or themes (ex. minor, major etc.)
- Compare ideas across stories, poems, or drama
- Infer authors purpose in multimedia (Explicit =add one egg, Inferred – use raw egg, crack it open)
- Manipulate information in correct sequential order
- Use correct subject/verb agreement in written or spoken language
- Apply the correct ending punctuation
- Compare two texts to determine fact or opinion

Level 3: Students performing at this level have a **solid command** of the knowledge and skills contained in Extended Content Standards of the *English Language Arts Common Core State Standards (CCSS)* at Grade 6 and are academically prepared to engage successfully in further studies in this content area.

A student performing at Grade 6 will use adapted age appropriate materials to complete tasks including:

- Describe characters actions in a story
- Describe what the narrator or speaker in a story is thinking or feeling
- Determine events or actions that are stated explicitly (add one egg)
- Determine sequential order from informational text
- Label parts of speech (ex. Nouns, verbs, adjectives) in written text
- Select the correct ending punctuation to a sentence
- Determine fact and opinion statements in text

Level 2: Students performing at this level have a **partial command** of the knowledge and skills contained in the Extended Content Standards of the *English Language Arts Common Core State Standards (CCSS)* at Grade 6 and *will likely need academic support to engage successfully* in further studies in this content area.

A student performing at Grade 6 will use adapted age appropriate materials to complete tasks including:

- Identify specific characters in a story
- Identify (via picture and written text) the narrator or speakers feelings
- Identify examples and anecdotes that relate to key individuals, events, or ideas in a text
- Determine the beginning and end of an action or event using informational text (add one egg and cook)
- Identify examples which represent parts of speech
- Identify the different ending punctuation marks in sentences
- Identify similar events across texts

Level 1: Students performing at this level have a **limited command** of the knowledge and skills contained in the Extended Content Standards of the *English Language Arts Common Core State Standards (CCSS)* at Grade 6 and *will need academic support to engage successfully* in further studies in this content area.

A student performing at Grade 6 will use adapted age appropriate materials to complete tasks including:

- Identify what a character is in a story
- Identify words in a story or written text
- Identify examples which represent key ideas in the text (ex: point to the egg)
- Identify sequential order words (ex. First, second, third)
- Identify the symbolic representation to the written or spoken word
- Identify a capital letter at the start of a sentence or proper noun. (David)
- Identify a fact (ex. Ball is round)

ELA Grade 7

Level 4: Students performing at this level have a **superior command** of the knowledge and skills contained in the Extended Content Standards of the *English Language Arts Common Core State Standards (CCSS)* in Grade 7 and are academically *well prepared to engage successfully* in further studies in this content area.

A student performing at Grade 7 will use adapted age appropriate materials to complete tasks including:

- Describe the characters, setting, and theme
- Determine what words an author uses to contrast characters in a text
- Determine how two or more events in a text are related (cause and effect)
- Write a narrative about persona or imagined experience or events from beginning to end
- Combine two simple sentences using common conjunctions to produce compound sentences

Level 3: Students performing at this level have a **solid command** of the knowledge and skills contained in Extended Content Standards of the *English Language Arts Common Core State Standards (CCSS)* at Grade 7 and are academically prepared to engage successfully in further studies in this content area.

A student performing at Grade 7 will use adapted age appropriate materials to complete tasks including:

- Compare how two or more characters relate to each other
- Determine whether a text is a story, drama, or poem
- Determine two or more central ideas in a text
- Use words or phrases to describe characters or events
- Produce simple sentences using capitalization and ending punctuation

Level 2: Students performing at this level have a **partial command** of the knowledge and skills contained in the Extended Content Standards of the *English Language Arts Common Core State Standards (CCSS)* at Grade 7 and *will likely need academic support to engage successfully* in further studies in this content area.

A student performing at Grade 7 will use adapted age appropriate materials to complete tasks including:

- Identify explicit character traits in a story
- Identify the difference between a story and a poem
- Identify one explicit statement (ex: Animals eat plants to live)
- Use words to describe one or more characters
- Can produce simple sentences

Level 1: Students performing at this level have a **limited command** of the knowledge and skills contained in the Extended Content Standards of the *English Language Arts Common Core State Standards (CCSS)* at Grade 7 and *will need academic support to engage successfully* in further studies in this content area.

A student performing at Grade 7 will use adapted age appropriate materials to complete tasks including:

- Identify specific characters and setting in a story
- Identify rhyme and repetition of sounds in a text
- Identify the central idea of the text (ex. The brown bear)
- Use words to signal event order
- Identify which picture represents the correct sentence (ex. The ball is round)

ELA Grade 8

Level 4: Students performing at this level have a **superior command** of the knowledge and skills contained in the Extended Content Standards of the *English Language Arts Common Core State Standards (CCSS)* in Grade 8 and are academically *well prepared to engage successfully* in further studies in this content area.

A student performing at Grade 8 will use adapted age appropriate materials to complete tasks including:

- Use supporting details to back up a statement (inference, theme, cause and effect relationship)
- Use word analysis and story structure to create meaning and evidence throughout various types of media
- Compare multiple text on the same topic to identify conflicting evidence
- Distinguish between fact and opinion and provide evidence to support

Level 3: Students performing at this level have a **solid command** of the knowledge and skills contained in Extended Content Standards of the *English Language Arts Common Core State Standards (CCSS)* at Grade 8 and are academically prepared to engage successfully in further studies in this content area.

A student performing at Grade 8 will use adapted age appropriate materials to complete tasks including:

- Summarize theme using supporting details
- Determine the meaning of informational words using context clues
- Determine patterns, events, or characters within the text
- Compare fact and opinion

Level 2: Students performing at this level have a **partial command** of the knowledge and skills contained in the Extended Content Standards of the *English Language Arts Common Core State Standards (CCSS)* at Grade 8 and *will likely need academic support to engage successfully* in further studies in this content area.

A student performing at Grade 8 will use adapted age appropriate materials to complete tasks including:

- Determine the central ideas and theme of text
- Determine the authors point of view
- Describe a pattern of a text
- Given a statement determine if it is a fact or opinion

Level 1: Students performing at this level have a **limited command** of the knowledge and skills contained in the Extended Content Standards of the *English Language Arts Common Core State Standards (CCSS)* at Grade 8 and *will need academic support to engage successfully* in further studies in this content area.

A student performing at Grade 8 will use adapted age appropriate materials to complete tasks including:

- Identify a detail of the text
- Identify the topic of the story

- Identify text similarities
- Identify a fact and provide evidence to support facts

ELA Grade 10

Level 4: Students performing at this level have a **superior command** of the knowledge and skills contained in the Extended Content Standards of the *English Language Arts Common Core State Standards (CCSS)* in Grade 10 and are academically *well prepared to engage successfully* in further studies in this content area.

- Determine which quote best demonstrates the meaning of the text or an inference drawn from a text.
- Compare and contrast the experience of characters with personal experience
- Determine which word in an array of content related words is missing from a sentence
- Determine the meaning of a word with multiple meanings in a text
- Correctly use commas in a sentence or letter
- Spell high frequency words correctly and use phonetic spelling for unknown words
- Correctly use capital letters for proper nouns in sentences
- Recognize that the story contains an altered sequence by identifying the beginning, middle, or end

Level 3: Students performing at this level have a **solid command** of the knowledge and skills contained in Extended Content Standards of the *English Language Arts Common Core State Standards (CCSS)* at Grade 10 and are academically prepared to engage successfully in further studies in this content area.

A student performing at Grade 10 will use adapted age appropriate materials to complete tasks including:

- Tell theme or central idea of a story
- Retell a story in proper sequence
- Answer inferential questions based on a text
- Identify a word or sentence that tells an author's point of view
- Support an answer using details from the story
- Use correct punctuation at the end of a sentence
- Use capitalization for beginning of sentences
- Determine fact/opinion
- Spell high frequency words correctly

Level 2: Students performing at this level have a **partial command** of the knowledge and skills contained in the Extended Content Standards of the *English Language Arts Common Core State Standards (CCSS)* at Grade 10 and *will likely need academic support to engage successfully* in further studies in this content area.

- Answer literal questions based on a given text
- Identify beginning, middle, and end in a story
- Identify the correct high frequency word when presented with an array of high frequency words
- Correctly use a period or question mark at the end of a sentence
- Identify a fact from the passage

Level 1: Students performing at this level have a **limited command** of the knowledge and skills contained in the Extended Content Standards of the *English Language Arts Common Core State Standards (CCSS)* at Grade 10 and *will need academic support to engage successfully* in further studies in this content area.

- Identify a character or an event from a story
- Identify a graphic that displays a scene from a story
- Recognize that a period goes at the end of a sentence
- Identify beginning letter or sounds from high frequency words

ELA Grade 11

Level 4: Students performing at this level have a **superior command** of the knowledge and skills contained in the Extended Content Standards of the *English Language Arts Common Core State Standards (CCSS)* in Grade 10 and are academically *well prepared to engage successfully* in further studies in this content area.

- Determine which quote best demonstrates the meaning of the text or an inference drawn from a text.
- Compare and contrast the experience of characters with personal experience
- Determine which word in an array of content related words is missing from a sentence
- Determine the meaning of a word with multiple meanings in a text
- Correctly use commas in a sentence or letter
- Spell high frequency words correctly and use phonetic spelling for unknown words
- Correctly use capital letters for proper nouns in sentences
- Recognize that the story contains an altered sequence by identifying the beginning, middle, or end

Level 3: Students performing at this level have a **solid command** of the knowledge and skills contained in Extended Content Standards of the *English Language Arts Common Core State Standards (CCSS)* at Grade 10 and are academically prepared to engage successfully in further studies in this content area.

A student performing at Grade 10 will use adapted age appropriate materials to complete tasks including:

- Tell theme or central idea of a story
- Retell a story in proper sequence
- Answer inferential questions based on a text
- Identify a word or sentence that tells an author's point of view
- Support an answer using details from the story
- Use correct punctuation at the end of a sentence
- Use capitalization for beginning of sentences
- Determine fact/opinion
- Spell high frequency words correctly

Level 2: Students performing at this level have a **partial command** of the knowledge and skills contained in the Extended Content Standards of the *English Language Arts Common Core State Standards (CCSS)* at Grade 10 and *will likely need academic support to engage successfully* in further studies in this content area.

- Answer literal questions based on a given text
- Identify beginning, middle, and end in a story
- Identify the correct high frequency word when presented with an array of high frequency words
- Correctly use a period or question mark at the end of a sentence
- Identify a fact from the passage

Level 1: Students performing at this level have a **limited command** of the knowledge and skills contained in the Extended Content Standards of the *English Language Arts Common Core State Standards (CCSS)* at Grade 10 and *will need academic support to engage successfully* in further studies in this content area.

- Identify a character or an event from a story
- Identify a graphic that displays a scene from a story
- Recognize that a period goes at the end of a sentence
- Identify beginning letter or sounds from high frequency words

Mathematics

Math Grade 3

Level 4: Students performing at this level have a **superior command** of the knowledge and skills contained in the Extended Content Standards of the *Mathematics Common Core State Standards (CCSS)* at Math Grade 3 and are academically *well prepared to engage successfully* in further studies in this content area.

Students will consistently demonstrate the ability to:

- Partition objects into equal halves and match to a fractional term
- Recognize attributes of a rhombus and other quadrilaterals
- Solve problems with length using appropriate vocabulary
- Label a line graph with title and axes
- Identify lengths of objects to nearest inch using a standard measurement tool
- Compose and decompose numbers on both sides of equal sign (e.g. 22 is the same as two tens and two ones)
- Use part-part-whole relationships to compose and decompose numbers 0-30
- Use manipulatives to solve real life problems using repeated additions and equal shares

Level 3: Students performing at this level have a **solid command** of the knowledge and skills contained in Extended Content Standards of the *Mathematics Common Core State Standards (CCSS)* at Math Grade 3 and are academically prepared to engage successfully in further studies in this content area.

Students will consistently demonstrate the ability to:

- Identify a whole or half of an object or shape
- Categorize shapes by number of sides
- Compare lengths of two objects using language (e.g. longer or shorter)
- Use a line plot and to answer basic questions (e.g. more, less, equal)
- Add/subtract using symbols up to 30 without regrouping
- Use manipulatives to build models to solve real life problems involving equal groups
- Compare numbers 0-30 using number line (e.g. greater than, less than)

Level 2: Students performing at this level have a **partial command** of the knowledge and skills contained in the Extended Content Standards of the *Mathematics Common Core State Standards (CCSS)* in Math Grade 3 and *will likely need academic support to engage successfully* in further studies in this content area.

Students will consistently demonstrate the ability to:

- Identify whole objects
- Identify basic shapes
- Sort objects by length (e.g. longer, shorter)

- Given a graph, match appropriate data
- Combine and take away from sets to tell how many
- Identify numbers on a number line up to 30

Level 1: Students performing at this level have a **limited command** of the knowledge and skills contained in the Extended Content Standards of the *Mathematics Common Core State Standards (CCSS)* in Math Grade 3 and *will need academic support to engage successfully* in further studies in this content area.

Students will consistently demonstrate the ability to:

- Match common shapes
- Match non-standard measurement objects to given stimulus
- Point to the graph upon request
- Match with 1-1 correspondence (object to object)
- Match corresponding numbers to a given number line (0-5)
- Given a model, create two equal sets
- Track numbers on a number line

Math Grade 4

Level 4: Students performing at this level have a **superior command** of the knowledge and skills contained in the Extended Content Standards of the *Mathematics Common Core State Standards (CCSS)* at Math Grade 4 and are academically *well prepared to engage successfully* in further studies in this content area.

Students will consistently demonstrate the ability to:

- Write a number sentence when given symbolic representation of a times or divide problem
- Use repeating shape patterns to make predictions
- Given numbers 1-50, identify the place value of each digit
- Use a numberline to identify the half between each number
- Identify the shape given specific attributes
- Solve problems using appropriate vocabulary to describe difference in weight (e.g. more, less, same)

Level 3: Students performing at this level have a **solid command** of the knowledge and skills contained in Extended Content Standards of the *Mathematics Common Core State Standards (CCSS)* at Math Grade 4 and are academically prepared to engage successfully in further studies in this content area.

Students will consistently demonstrate the ability to:

- Solve addition and subtraction problems when change is unknown (e.g. 8 + _ = 12)
- Identify correct number sentence when given a symbolic representation of multiplication or division
- Extend repeating pattern using shapes
- Illustrate whole numbers to 50 by composing and decomposing numbers
- Identify whole, half and fourth using concrete models and using symbolic representation
- Tell time to nearest hour
- Compare two objects using mass and weight (ounces, lbs)

Level 2: Students performing at this level have a **partial command** of the knowledge and skills contained in the Extended Content Standards of the *Mathematics Common Core State Standards (CCSS)* in Math Grade 4 and *will likely need academic support to engage successfully* in further studies in this content area.

Students will consistently demonstrate the ability to:

- Using manipulatives, solve addition and subtraction problems to find answer
- Illustrate multiplication and division by making equal sized groups using models
- Complete A B pattern using shapes
- Using numberline or hundred chart, compare 2 numbers < > or =
- Identify whole and half using concrete models
- Identify angles in each shape
- Compare weights of objects using vocabulary (lighter or heavier)

Level 1: Students performing at this level have a **limited command** of the knowledge and skills contained in the Extended Content Standards of the *Mathematics Common Core State Standards (CCSS)* in Math Grade 4 and *will need academic support to engage successfully* in further studies in this content area.

Students will consistently demonstrate the ability to:

- Count sets to show how many
- Create equal groups by using 1-1 correspondence
- Match shapes to a given pattern
- Locate numbers on a hundred chart or number line
- Match whole, half and quarter to given representations
- Identify which shape has an angle
- Identify parts of a clock
- Identify tools used to measure weight of object

Math Grade 5

Level 4: Students performing at this level have a **superior command** of the knowledge and skills contained in the Extended Content Standards of the *Mathematics Common Core State Standards (CCSS)* at Math Grade 5 and are academically *well prepared to engage successfully* in further studies in this content area.

Students will consistently demonstrate the ability to:

- Illustrate and solve a number problem based on a real world situation
- Create shape and numerical patterns
- Manipulate whole numbers in groups of 1s and 10s by composing and decomposing
- Solve addition and subtraction problems when initial is unknown
- Add fractions with like denominators to make a whole (halves, thirds, fourths)
- Compare the weight and length of an object using two different units (standard and nonstandard)
- Identify more, less and same on graphs

Level 3: Students performing at this level have a **solid command** of the knowledge and skills contained in Extended Content Standards of the *Mathematics Common Core State Standards (CCSS)* at Math Grade 5 and are academically prepared to engage successfully in further studies in this content area.

Students will consistently demonstrate the ability to:

- Solve a number problem based on a real world situation using addition and subtraction
- Demonstrate the concept of counting by 2s, 5s and 10s with numbers 0-100
- Correctly order counting numbers 0-100
- Solve single and multi-digit addition and subtraction equations with no regrouping
- Identify whole, half and fourth using concrete models
- Tell time to the nearest five minutes
- Display data on a picture of bar graph given two pieces of data
- Sort geometric figures based on common attributes

Level 2: Students performing at this level have a **partial command** of the knowledge and skills contained in the Extended Content Standards of the *Mathematics Common Core State Standards (CCSS)* in Math Grade 5 and *will likely need academic support to engage successfully* in further studies in this content area.

Students will consistently demonstrate the ability to:

- Identify the next shape or number in a given pattern
- Complete a pattern using manipulatives to count by 10s
- Correctly order numbers 0-50
- Solve single digit addition and subtraction equations with manipulatives
- Identify whole, half and fourth using concrete models
- Tell time to the nearest hour and half-hour

- Answer basic questions using a picture graph focusing on more, less, same
- Recognize basic shapes with common attributes

Level 1: Students performing at this level have a **limited command** of the knowledge and skills contained in the Extended Content Standards of the *Mathematics Common Core State Standards (CCSS)* in Math Grade 5 and *will need academic support to engage successfully* in further studies in this content area.

Students will consistently demonstrate the ability to:

- Match concrete objects to a pre-made pattern
- Correctly match numbers to each other (0-20)
- Solve single digit addition equations using manipulatives or a number line
- Create a whole object when given parts (halves, fourths and thirds)
- Identify most and least on a graph
- Match basic shapes with common attributes

Math Grade 6

Level 4: Students performing at this level have a **superior command** of the knowledge and skills contained in the Extended Content Standards of the *Mathematics Common Core State Standards (CCSS)* in 6th Grade Math and are academically *well prepared to engage successfully* in further studies in this content area.

Students will consistently demonstrate the ability to:

- Create ratios to represent relationships between 2 quantities
- Solve addition of fractions with like denominators to make a whole number
- Solve multiplication problems using numbers from 0-10
- Justify the answer for expressions
- Determine the area of rectangular figures using rows and provided columns
- Summarize and interpret data from a chart or a graph

Level 3: Students performing at this level have a **solid command** of the knowledge and skills contained in Extended Content Standards of the *Mathematics Common Core State Standards (CCSS)* at 6th Grade Math and are academically prepared to engage successfully in further studies in this content area.

Students will consistently demonstrate the ability to:

- Compare part-part and part-whole relationship
- Compare unit fractions
- Add fractions with like denominators
- Evaluate expressions for the variable using addition and subtraction
- Determine the perimeter of rectangular figures using given dimensions
- Display data in chart/graph (bar, picture, line plots)

Level 2: Students performing at this level have a **partial command** of the knowledge and skills contained in the Extended Content Standards of the *Mathematics Common Core State Standards (CCSS)* in 6th Grade Math and *will likely need academic support to engage successfully* in further studies in this content area.

- Compare part to whole
- Compare whole rational numbers
- Choose operation needed to solve expressions
- Distinguish the difference between area and perimeter
- Conduct surveys

Level 1: Students performing at this level have a **limited command** of the knowledge and skills contained in the Extended Content Standards of the *Mathematics Common Core State Standards (CCSS)* in 6th Grade Math and *will need academic support to engage successfully* in further studies in this content area.

- Recognize whole numbers

- Compare positive whole numbers
- Identify expressions and equations
- Identify corresponding (opposite) sides
- Select statistical questions

Math Grade 7

Level 4: Students performing at this level have a **superior command** of the knowledge and skills contained in the Extended Content Standards of the *Mathematics Common Core State Standards (CCSS)* in 7th Grade Math and are academically *well prepared to engage successfully* in further studies in this content area.

Students will consistently demonstrate the ability to:

- Apply equivalent ratios to solve problems
- Use all operations to solve problems with whole numbers 0-100 and greater and justify answer
- Subtract fractions with like denominators within fraction families (1/2, 1/3, ¼, 1/5, 1/6, 1/8, 1/10) with fraction bars
- Apply the properties of operations and equality to solve problems for unknown quantities
- Solve real life mathematical problems to find area
- Investigate chance process and develop, use, and evaluate probability models
- Draw informal comparative inferences about two populations

Level 3: Students performing at this level have a **solid command** of the knowledge and skills contained in Extended Content Standards of the *Mathematics Common Core State Standards (CCSS)* at 7th Grade Math and are academically prepared to engage successfully in further studies in this content area.

Students will consistently demonstrate the ability to:

- Model equivalent ratios
- Use all operations to solve problems with whole numbers 0-100
- Subtract fractions with like denominators using fraction families (1/8, 1/10) with fraction bars
- Use addition/multiplication properties to identify and illustrate equivalent expressions
- Use multiplication to solve area of rectangles
- Use survey data to interpret and compare data from two graphs
- Determine the probability of an event being possible or impossible

Level 2: Students performing at this level have a **partial command** of the knowledge and skills contained in the Extended Content Standards of the *Mathematics Common Core State Standards (CCSS)* in 7th Grade Math and *will likely need academic support to engage successfully* in further studies in this content area.

- Identify equivalent ratios
- Compute using 1 digit and 2 digit whole numbers
- Subtract fractions with like denominators using fraction families (1/5, 1/6) with fraction bars
- Define properties (associative and commutative)
- Use area formula (L x W = A)
- Identify representative random sample
- Define probability

Level 1: Students performing at this level have a **limited command** of the knowledge and skills contained in the Extended Content Standards of the *Mathematics Common Core State Standards (CCSS)* in 7th Grade Math and *will need academic support to engage successfully* in further studies in this content area.

- Create ratios to represent relationships between 2 quantities
- Add, subtract, multiply, divide whole numbers 0-10
- Subtract fractions with like denominators using fraction families (1/2, 1/3, ¼) using fraction bars
- Demonstrate that the sum of zero and a number stays the same value
- Using knowledge of rows and columns, identify length and width of a rectangle
- Differentiate between a population and sample
- Define an event

Math Grade 8

Level 4: Students performing at this level have a **superior command** of the knowledge and skills contained in the Extended Content Standards of the *Mathematics Common Core State Standards (CCSS)* in 8th Grade Math and are academically *well prepared to engage successfully* in further studies in this content area.

Students will consistently demonstrate the ability to:

- Demonstrate the connections between proportional relationships and lines plotted
- Analyze and solve linear equations using whole numbers
- Use physical models with various orientations to find congruency (proximity, positions, directions, turns)
- Solve real world math problems involving volume of rectangular prisms
- Make predictions using trends in existing data from scatter plots

Level 3: Students performing at this level have a **solid command** of the knowledge and skills contained in Extended Content Standards of the *Mathematics Common Core State Standards (CCSS)* at 8th Grade Math and are academically prepared to engage successfully in further studies in this content area.

Students will consistently demonstrate the ability to:

- Find and graph equivalent ratios in 1st quadrant
- Use equations to solve problems involving whole numbers all operations when a part is unknown
- Determine the congruence of polygons with given attributes
- Measure volumes of right rectangular figures by counting unit cubes
- Analyze patterns in scatter plots to determine trends as positive, negative, or no association

Level 2: Students performing at this level have a **partial command** of the knowledge and skills contained in the Extended Content Standards of the *Mathematics Common Core State Standards (CCSS)* in 8th Grade Math and *will likely need academic support to engage successfully* in further studies in this content area.

- Graph points from a function table in 1st quadrant made from equivalent ratios
- Identify operation to find unknown part in an equation or expression
- Identify corresponding attributes of different figures
- Define volume a rectangular right prism
- Given data construct scatter plot
- Describe patterns found in a scatter plot

Level 1: Students performing at this level have a **limited command** of the knowledge and skills contained in the Extended Content Standards of the *Mathematics Common Core State Standards (CCSS)* in 8th Grade Math and *will need academic support to engage successfully* in further studies in this content area.

- Define unit rate (cost per unit)
- Identify 4 quadrants of coordinate plane
- Identify parts of equation

- Identify attributes of figures (faces, sides, angles)
- Identify right rectangular prism
- Define attributes of scatter plot

Math Grade 10

Level 4: Students performing at this level have a **superior command** of the knowledge and skills contained in the Extended Content Standards of the *Mathematics Common Core State Standards (CCSS)* in Math 1 and are academically *well prepared to engage successfully* in further studies in this content area.

Students will consistently demonstrate the ability to:

- Identify, order, add, and subtract decimals to the hundredths place to compare values or set up or solve equations
- Use inequality symbols to compare quantities or make inequalities true by replacing unknown variables with non-negative (e.g., 0, 1, 2, 3) integers
- Identify or interpret the unit rate (e.g., speed = mph) from a graph
- Use algebraic concepts to:
    - Identify equivalent expressions by combining like terms
    - Evaluate expressions by substituting numbers for variables
    - Set up or solve equations/inequalities using addition or subtraction in algebraic form or real life situations (e.g., word problems)

Level 3: Students performing at this level have a **solid command** of the knowledge and skills contained in Extended Content Standards of the *Mathematics Common Core State Standards (CCSS)* at Math 1 and are academically prepared to engage successfully in further studies in this content area.

Students will consistently demonstrate the ability to:

- Identify, order, add, or subtract decimals to compare values to the hundredths
- Use inequality terms to compare quantities (e.g., less than, smaller than)
- Identify positive integers that would make an inequality true (e.g., __ is greater than 7)
- Read a graph and identify quantities or units of measure
- Use algebraic concepts to:
    - Identify expressions by substituting numbers for variables
    - Solve one-step addition or subtraction equations with decimals or inequalities involving whole numbers for one unknown involving whole numbers for one unknown (e.g., 3.5 + __ = 5.5; 3 + __ is bigger than 5)

Level 2: Students performing at this level have a **partial command** of the knowledge and skills contained in the Extended Content Standards of the *Mathematics Common Core State Standards (CCSS)* in Math 1 and *will likely need academic support to engage successfully* in further studies in this content area.

- Identify, order, or compare decimals to tenths
- Determine which number is bigger and smaller
- Identify types of graphs (e.g., pie, bar) or subject of graph
- Use algebraic concepts to:
    - Solve one step addition equations without variables
    - Use variable to represent numbers (e.g., Let X = apples, there are 5 apples, what is X?)

Level 1: Students performing at this level have a **limited command** of the knowledge and skills contained in the Extended Content Standards of the *Mathematics Common Core State Standards (CCSS)* in Math 1 and *will need academic support to engage successfully* in further studies in this content area.

- Identify decimals through tenths place
- Determine which number is bigger or smaller
- Identify a graphic representation of data

Math Grade 11

Level 4: Students performing at this level have a **superior command** of the knowledge and skills contained in the Extended Content Standards of the *Mathematics Common Core State Standards (CCSS)* in Math 1 and are academically *well prepared to engage successfully* in further studies in this content area.

Students will consistently demonstrate the ability to:

- Identify, order, add, and subtract decimals to the hundredths place to compare values or set up or solve equations
- Use inequality symbols to compare quantities or make inequalities true by replacing unknown variables with non-negative (e.g., 0, 1, 2, 3) integers
- Identify or interpret the unit rate (e.g., speed = mph) from a graph
- Use algebraic concepts to:
    - Identify equivalent expressions by combining like terms
    - Evaluate expressions by substituting numbers for variables
    - Set up or solve equations/inequalities using addition or subtraction in algebraic form or real life situations (e.g., word problems)

Level 3: Students performing at this level have a **solid command** of the knowledge and skills contained in Extended Content Standards of the *Mathematics Common Core State Standards (CCSS)* at Math 1 and are academically prepared to engage successfully in further studies in this content area.

Students will consistently demonstrate the ability to:

- Identify, order, add, or subtract decimals to compare values to the hundredths
- Use inequality terms to compare quantities (e.g., less than, smaller than)
- Identify positive integers that would make an inequality true (e.g., __ is greater than 7)
- Read a graph and identify quantities or units of measure
- Use algebraic concepts to:
    - Identify expressions by substituting numbers for variables
    - Solve one-step addition or subtraction equations with decimals or inequalities involving whole numbers for one unknown involving whole numbers for one unknown (e.g., 3.5 + __ = 5.5; 3 + __ is bigger than 5)

Level 2: Students performing at this level have a **partial command** of the knowledge and skills contained in the Extended Content Standards of the *Mathematics Common Core State Standards (CCSS)* in Math 1 and *will likely need academic support to engage successfully* in further studies in this content area.

- Identify, order, or compare decimals to tenths
- Determine which number is bigger and smaller
- Identify types of graphs (e.g., pie, bar) or subject of graph
- Use algebraic concepts to:

- Solve one step addition equations without variables
- Use variable to represent numbers (e.g., Let X = apples, there are 5 apples, what is X?)

Level 1: Students performing at this level have a **limited command** of the knowledge and skills contained in the Extended Content Standards of the *Mathematics Common Core State Standards (CCSS)* in Math 1 and *will need academic support to engage successfully* in further studies in this content area.

- Identify decimals through tenths place
- Determine which number is bigger or smaller
- Identify a graphic representation of data

Science

Science Grade 5

Level 4: Students performing at this level have a **superior command** of the knowledge and skills contained in the Extended Content Standards of the *North Carolina Science Essential Standards* at Grade 5 and are academically *well prepared to engage successfully* in further studies in this content area.

Students will consistently demonstrate the ability to:

- Predict and identify conditions that effect motion (e.g. ramp increases speed; weight reduces speed)
- Identify, compare and classify physical or chemical changes
- Classify changes in matter as physical (reversible) or chemical (irreversible)
- Describe elements of different types of weather
- Given internal/external body parts, explain the functions
- Describe how the environment/ecosystem supports plans and animals within the ecosystem

Level 3: Students performing at this level have a **solid command** of the knowledge and skills contained in Extended Content Standards of the *North Carolina Science Essential Standards* at Grade 5 and are academically *prepared to engage successfully* in further studies in this content area.

Students will consistently demonstrate the ability to:

- Identify objects that would make it easier or harder to push or pull other objects under certain conditions
- Recognize the causes that affect speed of objects under certain conditions (e.g. ramps, wheels, weight-force)
- Identify and classify physical or chemical changes in matter
- Recognize reversible/irreversible changes of matter (e.g. baking- irreversible, ice- reversible)
- Distinguish between examples of different types of weather
- Given examples of severe weather, choose the appropriate location/shelter
- Relate body parts to functions (e.g. mouth- eat, nose- smell, ears- hear, eyes- see)
- Identify the different ecosystems and the animals and plants within these systems

Level 2: Students performing at this level have a **partial command** of the knowledge and skills contained in the Extended Content Standards of the *North Carolina Science Essential Standards* at Grade 5 and *will likely need academic support to engage successfully* in further studies in this content area.

Students will consistently demonstrate the ability to:

- Sort objects that can be pushed or pulled
- Identify objectives within pictures that are fast or slow under certain conditions (e.g. ramps, wheels, weight)
- Sort physical or chemical changes in matter
- Identify examples of different types of weather (with pictures)
- Match weather to functional needs (e.g. clothing, shelter, safety)
- Identify parts of the body, including internal organs
- Match body parts to essential functions (e.g. eyes- see, ears- hear, mouth- eat, nose- smell)
- Sort or match animals/plants to the correct ecosystem

Level 1: Students performing at this level have a **limited command** of the knowledge and skills contained in the Extended Content Standards of the *North Carolina Science Essential Standards* at Grade 5 and *will need academic support to engage successfully* in further studies in this content area.

Students will consistently demonstrate the ability to:

- Demonstrate pushing and pulling using an object
- Participate in an activity that shows a chemical and physical change (e.g. use a switch to work a blender)
- Recognize a picture/symbol of types of weather
- Match appropriate clothing needs to weather with pictures and objects
- Identify basic body parts using his/her body
  Sort living/non-living objects or pictures

Science Grade 8

Level 4: Students performing at this level have a **superior command** of the knowledge and skills contained in the Extended Content Standards of the *North Carolina Science Essential Standards* at their grade level and are academically *well prepared to engage successfully* in further studies in this content area.

- Compare and contrast energy sources and the appropriate uses of energy
- Determine remainder weight needed for parts to be whole.
- Explain why monitoring the hydrosphere and stewardship of water impacts human health
- Convey how to eliminate and treat illness
- Relate interconnectedness between living things and the environment
- Organize and sequence a complex food web

Level 3: Students performing at this level have a **solid command** of the knowledge and skills contained in Extended Content Standards of the *North Carolina Science Essential Standards* at their grade level and are academically *prepared to engage successfully* in further studies in this content area.

- Determine the effect of energy use (electricity turns on a light, fire burns wood)
- Compare an object's weight to its parts and determine that they are equal
- Link bodies of water to the existence of life dependent upon water (human life is effected by water conservation and pollution)
- Describe the components and their roles in a simple food chain (consumer, producer, decomposer)
- Convey how to prevent germs from causing illness and infection

Level 2: Students performing at this level have a **partial command** of the knowledge and skills contained in the Extended Content Standards of the *North Carolina Science Essential Standards* at their grade level and *will likely need academic support to engage successfully* in further studies in this content area.

- Recognize that energy produces an effect
- Recognize size and shape does not determine weight (whiffle ball vs. grapefruit)
- Distinguish between saltwater, freshwater, and polluted water
- Establish an illness/infection can be caused by a germ
- Determine which environmental factors are required to sustain human life (air, food, water)

Level 1: Students performing at this level have a **limited command** of the knowledge and skills contained in the Extended Content Standards of the *North Carolina Science Essential Standards* at their grade level and *will need academic support to engage successfully* in further studies in this content area.

- Identify basic form of energy (sun, lamp, fire)
- Demonstrate that an object has weight
- Identify various bodies of water(river, ocean, lake, stream)
- Identify type of illness/infection
- Identify living vs. non-living in an environment

Science Grade 10 (Biology)

Level 4: Students performing at this level have a **superior command** of the knowledge and skills contained in the Extended Content Standards of the *North Carolina Science Essential Standards* at their grade level and are academically *well prepared to engage successfully* in further studies in this content area.

- Identify raw materials needed for photosynthesis (i.e., water, sunlight, carbon dioxide)
- Understand basic anatomy and know functions of 7 major human body parts (i.e., skin, heart, brain, lungs, stomach, eyes, ears)
- Differentiate between unicellular and multi-cellular organisms
- Provide examples of fruits, vegetables, and meats that people eat
- Describe the role of plants and animals in the flow of energy through the environment
- Analyze or create a simple food chain or food web
- Recognize ways in which living things compete with each other to get resources
- Give examples of how human activities have impacted the environment
- Suggest ways to preserve natural resources

Level 3: Students performing at this level have a **solid command** of the knowledge and skills contained in Extended Content Standards of the *North Carolina Science Essential Standards* at their grade level and are academically *prepared to engage successfully* in further studies in this content area.

- Recognize that plants make food using the process of photosynthesis
- Understand basic human anatomy and know the functions of 5 major human body parts (i.e., skin, brain, heart, lungs, stomach, eyes, ears)
- Identify that the cell is the basic unit of life and most living things are made of cells
- Infer that fruits, vegetables, and meats as types of food that provide energy for people
- Review a simple food chain and identify the role of plants and animals in the flow of energy through the environment
- Understand that living things compete with each other to get resources
- Indicate how human activities impact the environment

Level 2: Students performing at this level have a **partial command** of the knowledge and skills contained in the Extended Content Standards of the *North Carolina Science Essential Standards* at their grade level and *will likely need academic support to engage successfully* in further studies in this content area.

- Know that plants make their own food
- Understand basic human anatomy and know the functions of 3 major human body parts (i.e., skin, brain, heart, lungs, stomach, eyes, ears)
- Identify that living things are made up of cells
- Indicate that people require food for energy
- Recognize a simple food chain
- Relate examples of natural resources and pollution

Level 1: Students performing at this level have a **limited command** of the knowledge and skills contained in the Extended Content Standards of the *North Carolina Science Essential Standards* at their grade level and *will need academic support to engage successfully* in further studies in this content area.

- Know that plants are living organisms
- Point out major internal and external body parts without giving functions (i.e., skin, brain, heart, lungs, stomach, eyes, ears)
- Recognize that cells are alive
- Indicate that people need food
- Understand that some living things eat other living things to survive
- Define or provide an example of the term “natural resource”

Science Grade 11 (Life Science)

Level 4: Students performing at this level have a **superior command** of the knowledge and skills contained in the Extended Content Standards of the *North Carolina Science Essential Standards* at their grade level and are academically *well prepared to engage successfully* in further studies in this content area.

- Identify, report, and follow proper procedure to respond to common disasters and accidents
- Give examples of simple and serious injuries
- Provide a relevant personal medical history
- Apply and recognize the importance of personal hygiene, proper nutrition, and exercise
- Plan, shop for, and prepare a nutritious meal
- Demonstrate food safety and how to properly prepare and/or store food

Level 3: Students performing at this level have a **solid command** of the knowledge and skills contained in Extended Content Standards of the *North Carolina Science Essential Standards* at their grade level and are academically *prepared to engage successfully* in further studies in this content area.

- Identify, report, and/or respond to common disasters and accidents
- Recognize types of injuries and proper procedures for treatment
- Provide relevant personal information
- Understand and apply the importance of personal hygiene, proper nutrition, and exercise
- Plan and prepare a nutritious meal
- Demonstrate food safety methods and how to properly prepare and store food

Level 2: Students performing at this level have a **partial command** of the knowledge and skills contained in the Extended Content Standards of the *North Carolina Science Essential Standards* at their grade level and *will likely need academic support to engage successfully* in further studies in this content area.

- Know how to respond to common disasters and accidents
- Distinguish between simple and serious injuries
- Provide limited personal and medical information (e.g., name, but not address; first name, but not last name)
- Recognize the importance of personal hygiene, proper nutrition, or exercise
- Plan a meal
- Identify methods of properly preparing or storing food

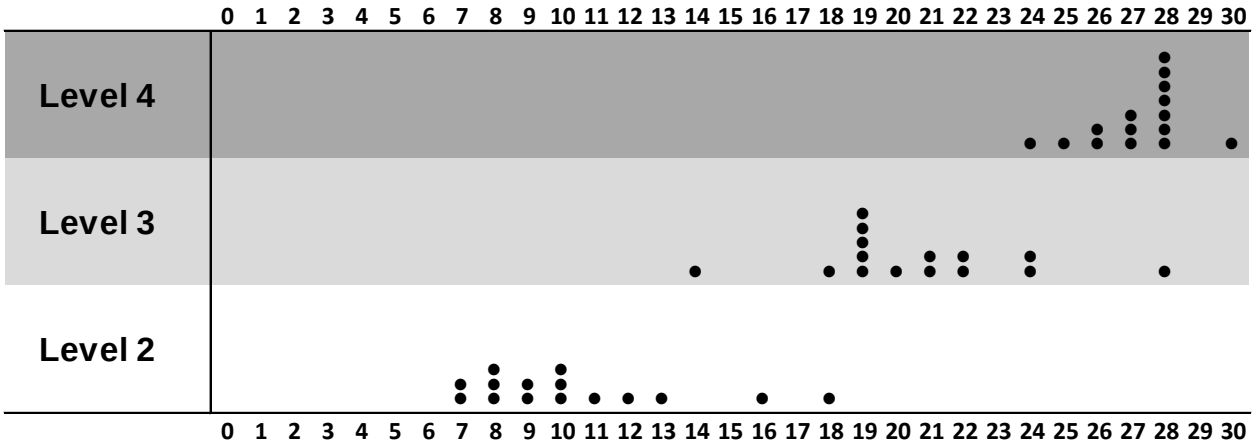
Level 1: Students performing at this level have a **limited command** of the knowledge and skills contained in the Extended Content Standards of the *North Carolina Science Essential Standards* at their grade level and *will need academic support to engage successfully* in further studies in this content area.

- Understand and recognize common disaster and accident procedures
- Identify between simple and serious injuries
- Provide minimum personal information (e.g., name)
- Define or provide an example of personal hygiene
- Recognize food
- Know that food needs to be stored properly

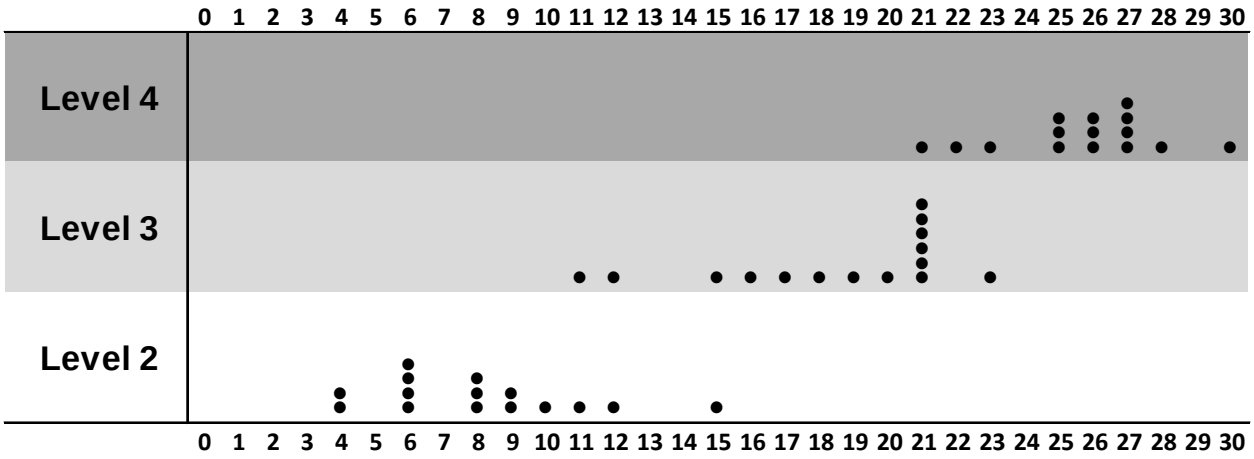
Appendix B: Graphical Display of Standard Setting Results by Subject and Grade

English Language Arts

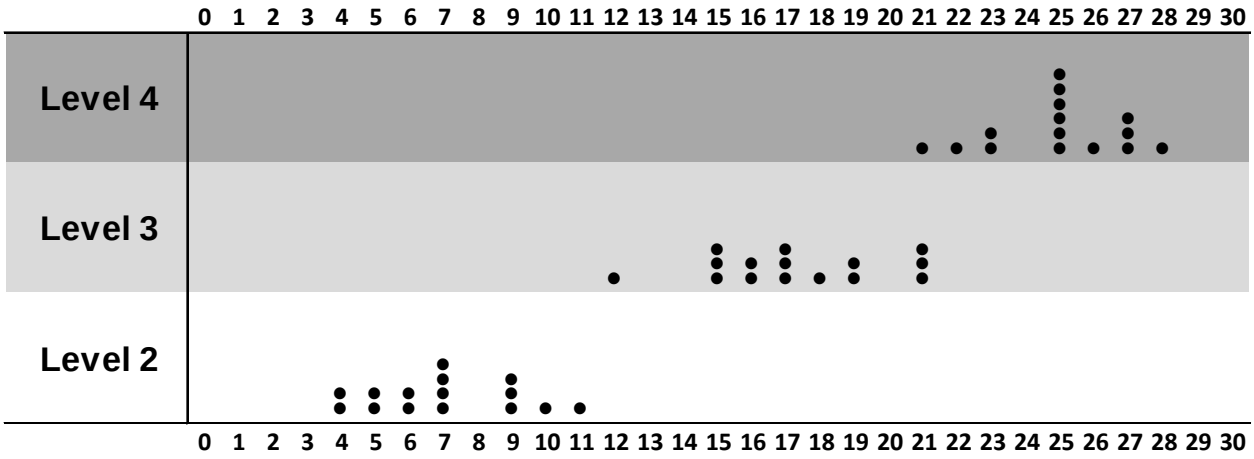
Grade 3



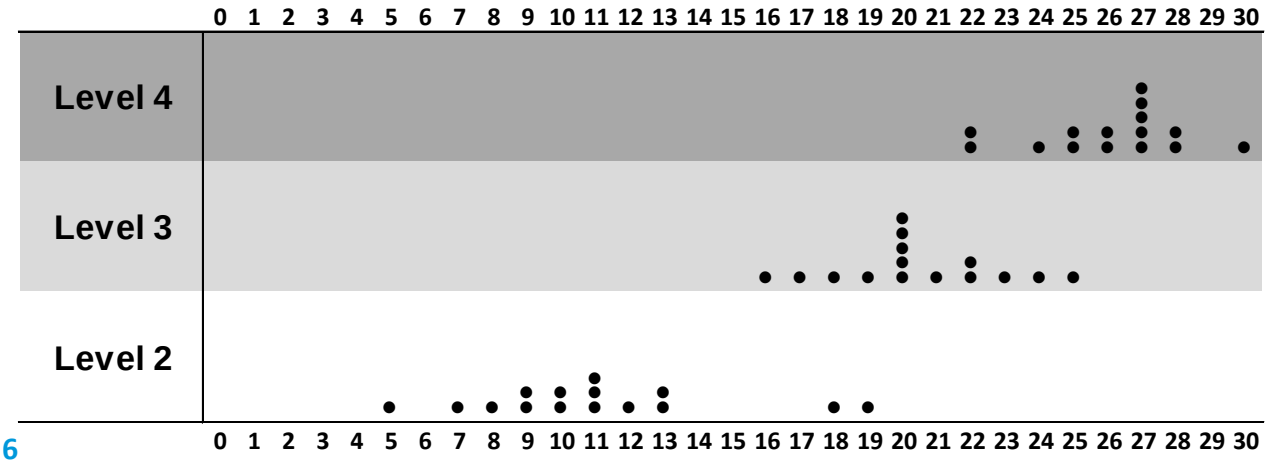
Grade 4



Grade 5

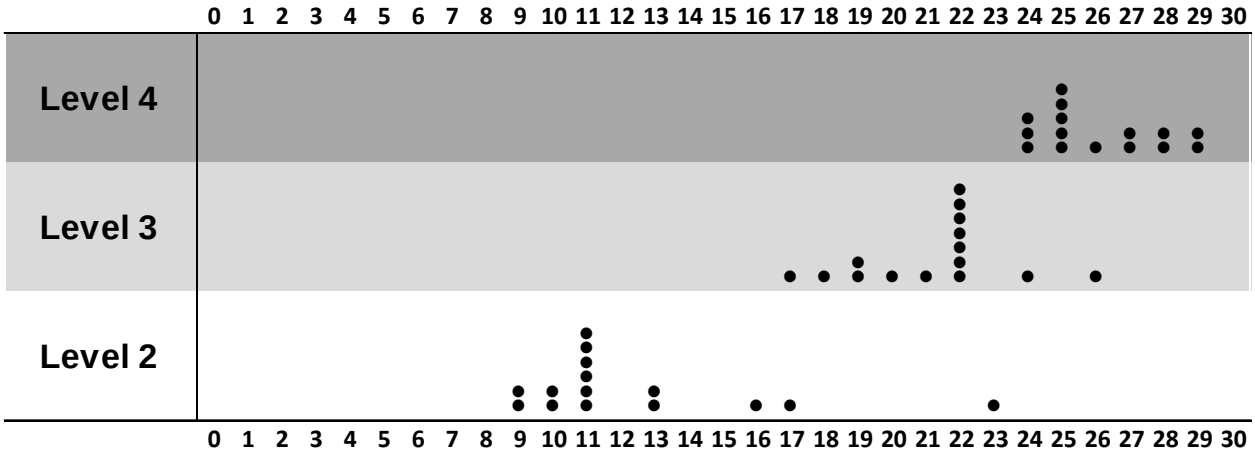


Grade

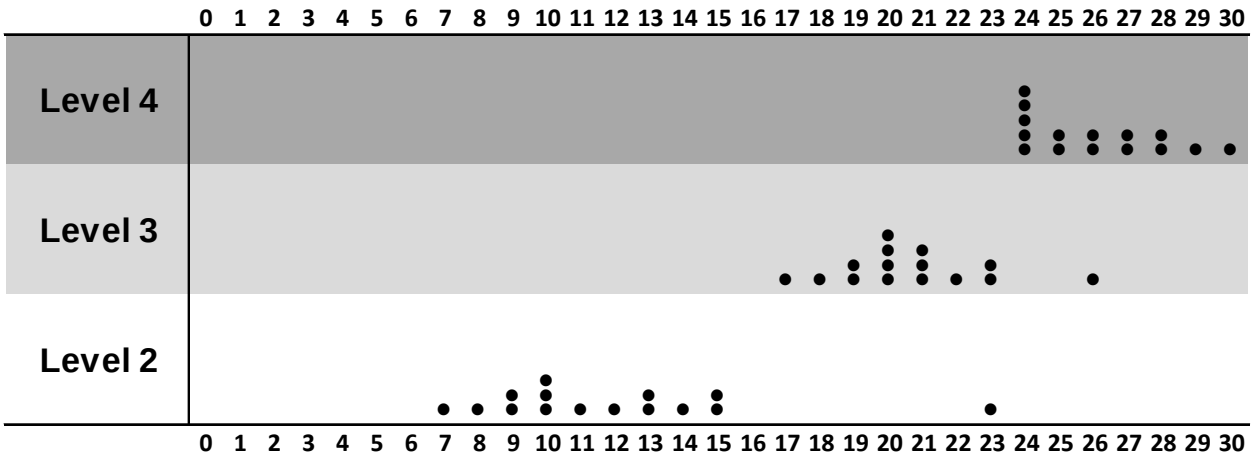


6

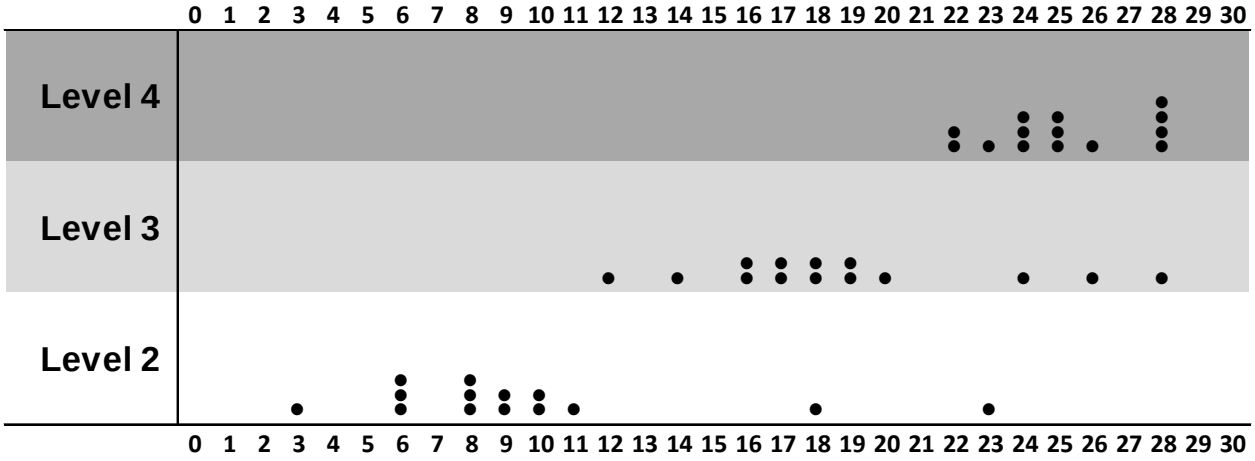
Grade 7



Grade 8

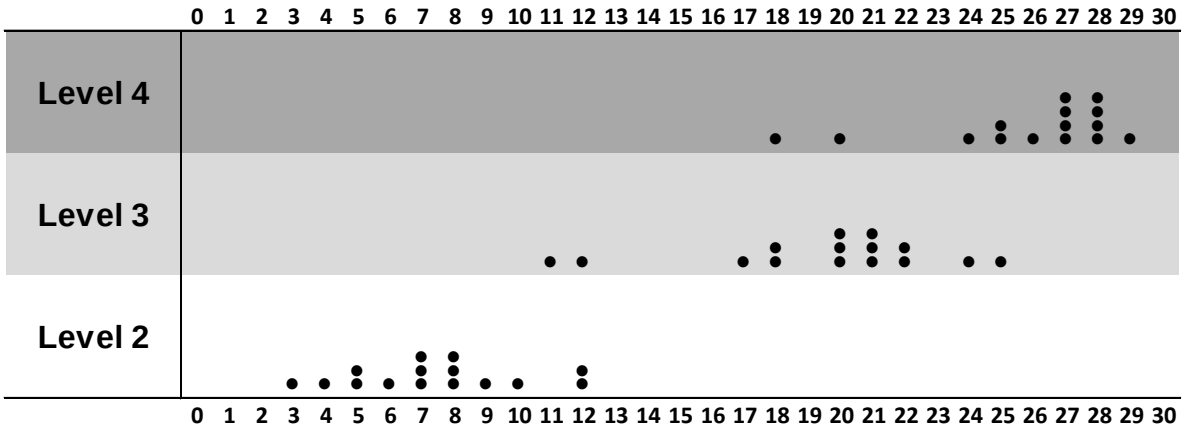


Grade 10/English II

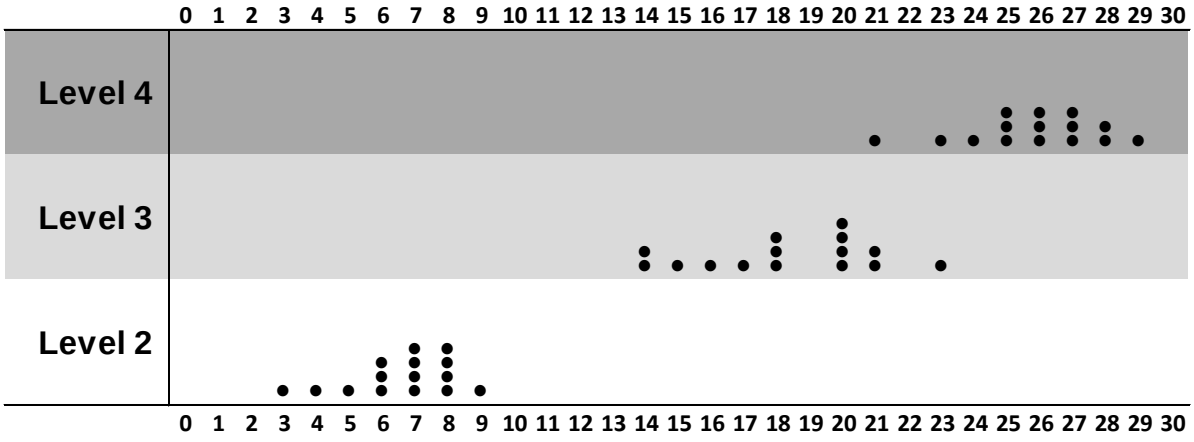


Mathematics

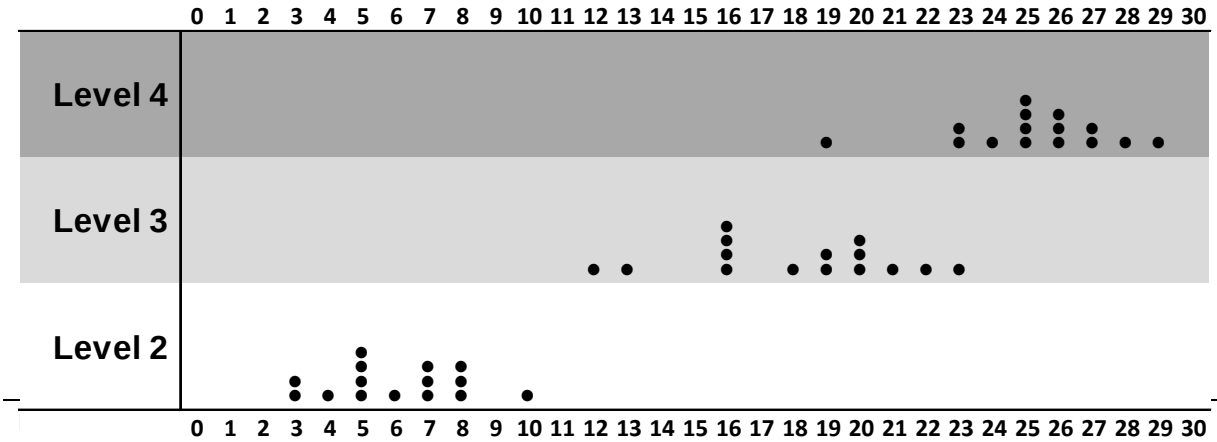
Grade 3



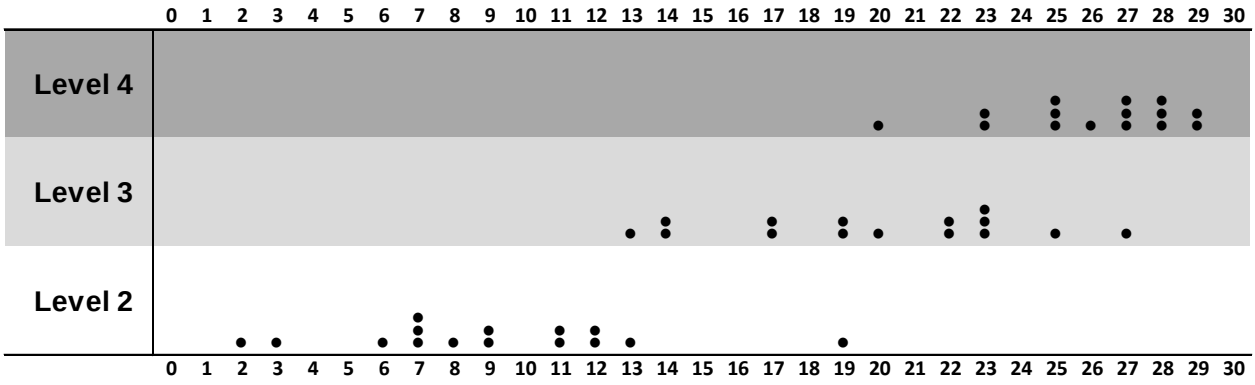
Grade 4



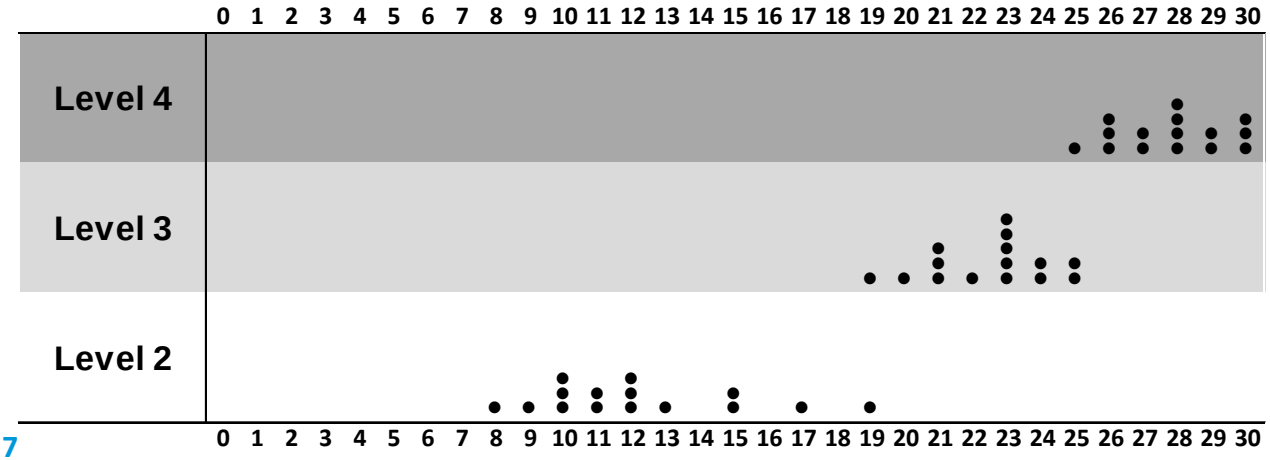
Grade 5



Grade 6

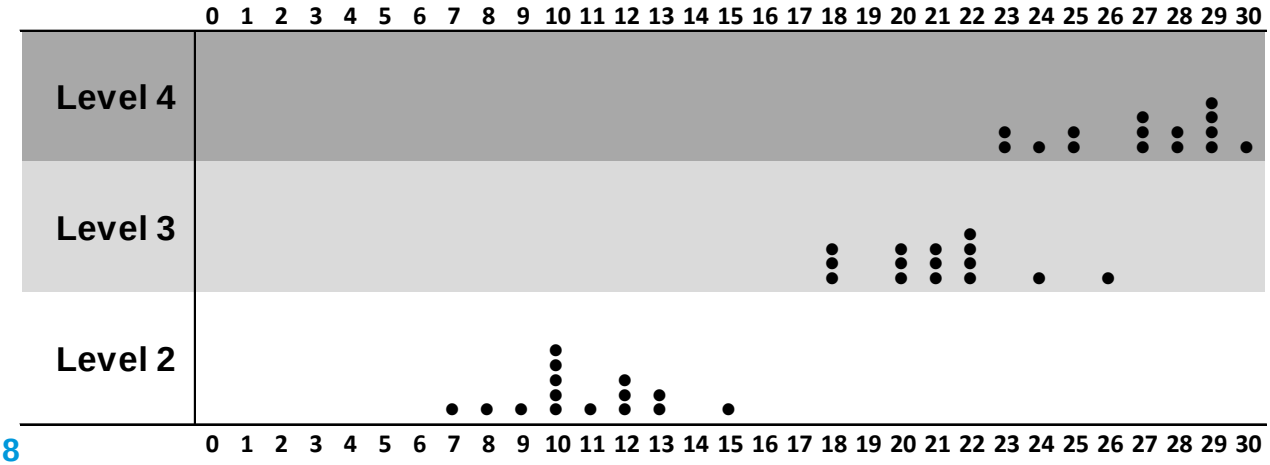


Grade



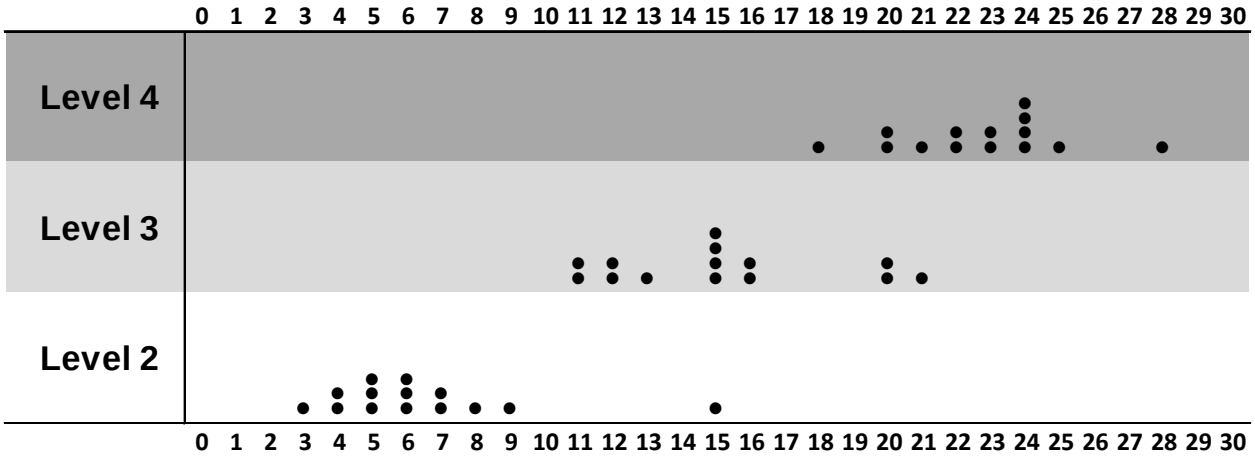
7

Grade



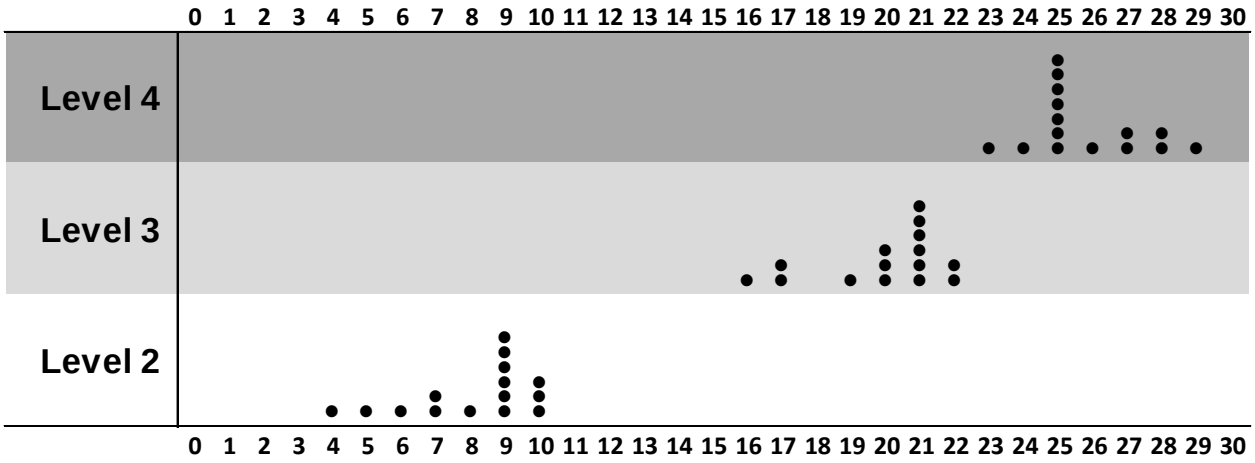
8

Grade 10/Algebra I

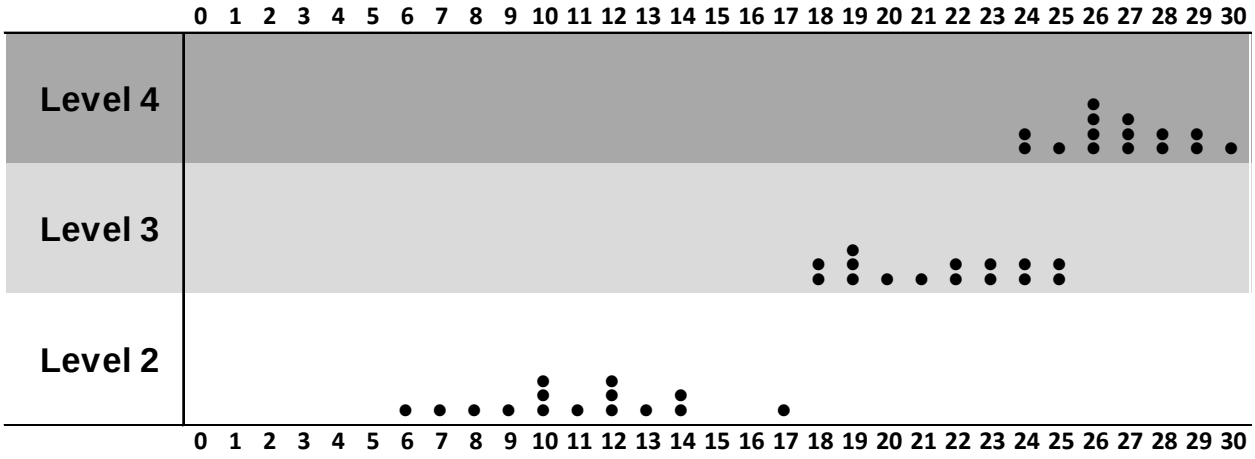


Science

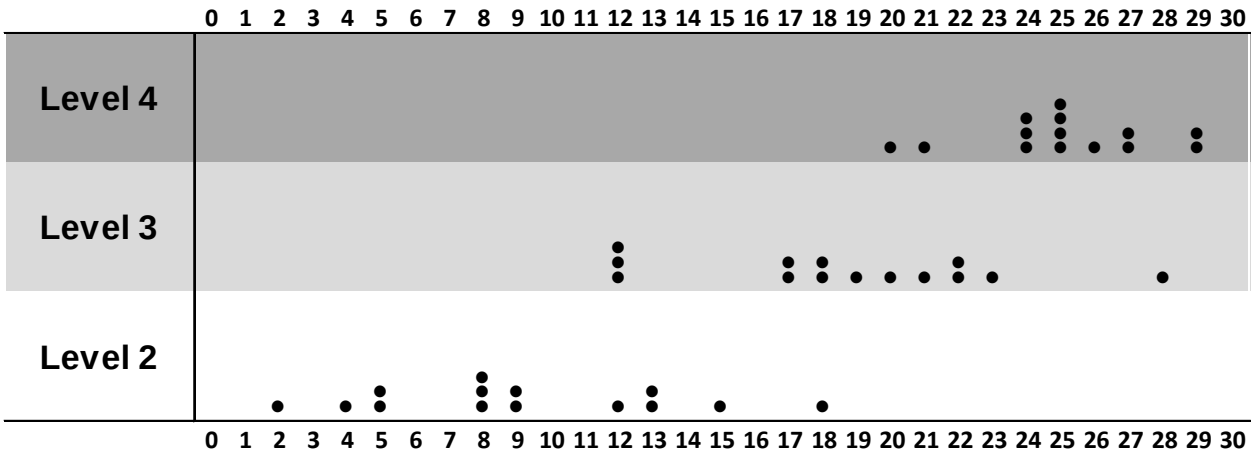
Grade 5



Grade 8

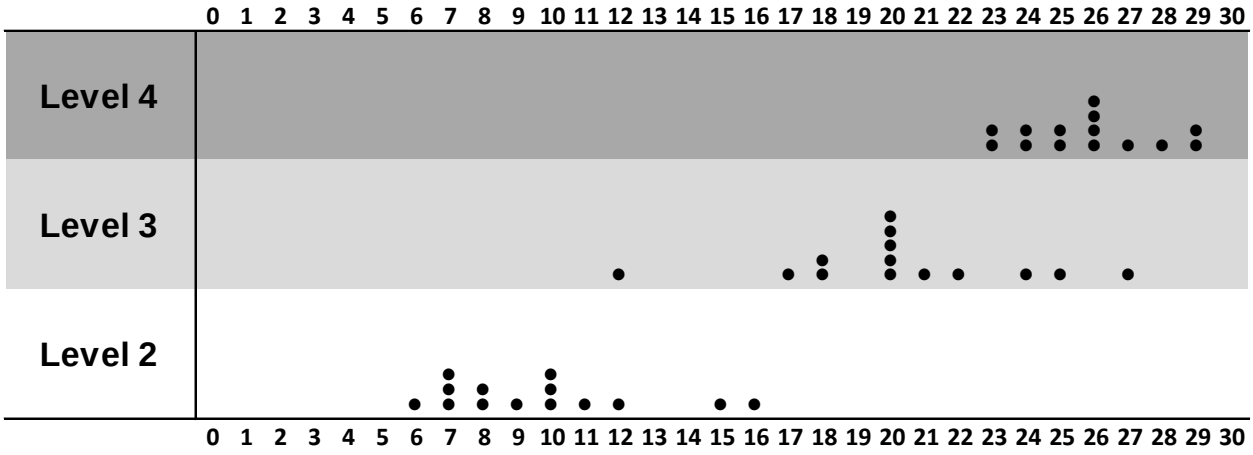


Grade 10/Biology



Mixed Subjects

Grade 11



Appendix C: Impact Tables by Subject Area

The tables in this appendix provide the estimated impact of any cut score. Specifically, the values in each table indicate what percent of students (at each grade level) scored at or above each possible score point based on the results of the 2012-2013 administration of the NCEXTEND1 tests.

English Language Arts

SCORE	3	4	5	6	7	8	10/English II
0	100%	100%	100%	100%	100%	100%	100%
1	98%	99%	98%	99%	98%	99%	98%
2	98%	99%	98%	99%	98%	99%	98%
3	97%	98%	98%	98%	97%	98%	98%
4	97%	98%	98%	98%	97%	98%	98%
5	97%	98%	97%	98%	97%	98%	98%
6	97%	98%	96%	98%	96%	97%	98%
7	96%	97%	96%	97%	96%	97%	97%
8	95%	97%	95%	97%	96%	97%	97%
9	95%	96%	95%	97%	96%	96%	96%
10	94%	96%	94%	96%	95%	95%	95%
11	93%	94%	93%	95%	94%	94%	93%
12	92%	93%	92%	94%	93%	92%	92%
13	91%	91%	90%	92%	91%	91%	90%
14	88%	88%	86%	89%	87%	88%	87%
15	84%	84%	82%	85%	82%	83%	82%
16	78%	80%	78%	80%	77%	78%	75%
17	74%	73%	72%	75%	72%	70%	69%
18	69%	68%	66%	69%	67%	65%	61%
19	64%	64%	60%	64%	62%	59%	55%
20	60%	60%	54%	58%	57%	53%	49%
21	56%	54%	48%	52%	53%	48%	44%
22	52%	48%	44%	47%	48%	43%	39%
23	49%	42%	41%	40%	43%	37%	35%
24	46%	36%	36%	34%	38%	31%	30%
25	43%	30%	29%	29%	34%	25%	26%
26	40%	26%	24%	23%	29%	18%	22%
27	34%	20%	19%	16%	22%	13%	16%
28	29%	12%	13%	11%	16%	9%	11%
29	22%	7%	8%	5%	10%	3%	7%

30	16%	2%	4%	1%	4%	1%	4%

Mathematics

SCORE	3	4	5	6	7	8	10/Algebra
0	100%	100%	100%	100%	100%	100%	100%
1	98%	99%	98%	99%	98%	99%	98%
2	98%	99%	98%	99%	98%	99%	98%
3	98%	98%	97%	98%	97%	98%	98%
4	97%	98%	97%	98%	97%	98%	97%
5	97%	97%	97%	98%	97%	98%	97%
6	97%	97%	96%	97%	97%	97%	96%
7	96%	97%	96%	97%	97%	97%	96%
8	95%	97%	96%	97%	96%	96%	96%
9	95%	97%	95%	96%	95%	96%	96%
10	94%	96%	95%	96%	95%	95%	95%
11	93%	95%	94%	95%	94%	94%	94%
12	91%	93%	92%	94%	93%	92%	92%
13	90%	92%	90%	91%	91%	89%	88%
14	86%	88%	85%	87%	88%	83%	85%
15	81%	83%	79%	82%	82%	78%	77%
16	75%	77%	72%	78%	76%	69%	68%
17	68%	69%	65%	70%	67%	61%	60%
18	60%	59%	58%	62%	60%	50%	51%
19	53%	52%	50%	53%	51%	40%	43%
20	46%	44%	43%	47%	44%	32%	35%
21	39%	38%	35%	40%	36%	26%	27%
22	33%	33%	28%	32%	29%	19%	21%
23	27%	27%	24%	25%	22%	14%	15%
24	22%	22%	19%	19%	17%	10%	11%
25	17%	17%	15%	14%	14%	7%	9%
26	13%	13%	10%	9%	10%	4%	6%
27	7%	11%	7%	5%	6%	2%	4%
28	4%	8%	4%	3%	4%	1%	2%
29	2%	5%	2%	1%	2%	1%	2%
30	1%	2%	1%	0%	1%	0%	1%

Science

SCORE	5	8	10/ Biology
0	100%	100%	100%
1	99%	98%	99%
2	98%	98%	98%
3	98%	98%	98%
4	97%	98%	98%
5	97%	97%	98%
6	97%	97%	98%
7	96%	97%	97%
8	96%	97%	97%
9	96%	97%	96%
10	96%	96%	96%
11	95%	96%	95%
12	94%	95%	94%
13	93%	94%	91%
14	91%	92%	89%
15	86%	89%	86%
16	82%	86%	80%
17	75%	80%	74%
18	71%	75%	68%
19	66%	70%	62%
20	60%	64%	56%
21	54%	58%	48%
22	49%	54%	40%
23	42%	50%	33%
24	37%	45%	27%
25	30%	39%	20%
26	25%	34%	16%
27	18%	28%	11%
28	12%	20%	8%
29	6%	12%	5%
30	2%	5%	2%

SCORE	11
0	100%
1	98%
2	98%
3	97%
4	97%
5	97%
6	96%
7	96%
8	96%
9	96%
10	95%
11	94%
12	93%
13	91%
14	89%
15	85%
16	79%
17	73%
18	66%
19	60%
20	54%
21	46%
22	41%
23	35%
24	29%
25	24%
26	20%
27	15%
28	9%
29	5%
30	2%

Mixed Subjects

Appendix D: Explanation of Guessing Adjustment

When test designs result in high probabilities of students meeting cut score thresholds by randomly guessing, it is reasonable to consider adjustments for guessing in the standard setting process (Cizek, 2012). The design of the NCEXTEND1 Assessments makes achieving high scores by guessing very likely (see Figure D1). For example, a student has an approximately 26% chance of scoring 20 or higher if the student randomly guessed on all questions. The effect of guessing is even more pronounced at lower score points: a student has an approximately 80% chance of scoring 10 or higher when guessing.

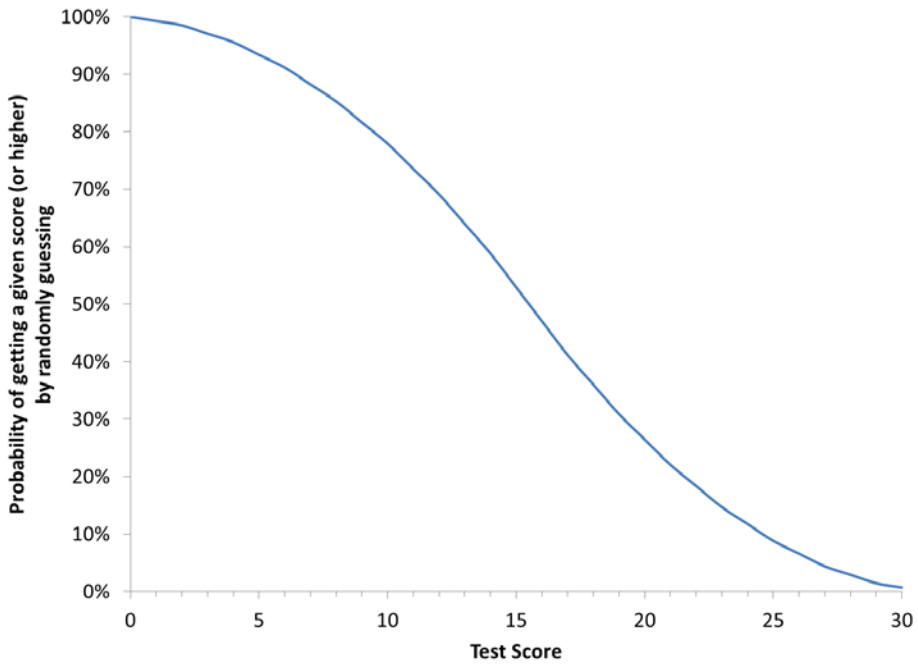
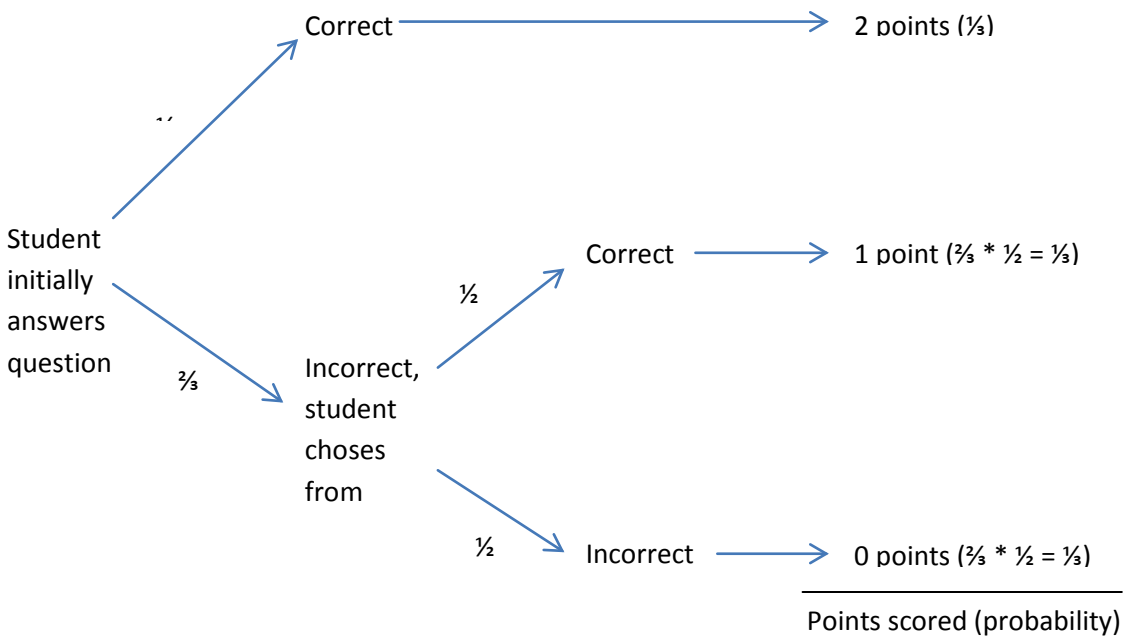


Figure D1. Probability of getting a test score (or higher) due to guessing, given the North Carolina Extend1 Assessment's scoring rules

The two-tiered scoring process for the EXTEND1 Assessments makes identifying a clear-cut and defensible guessing adjustment difficult. However, probabilistic theory can be applied to the scoring design, which can inform the decisions as to how to make cut score adjustments. One important probabilistic feature of the scoring design is that when a student guesses on an item, his/her expected score for that item is 1 (see Figure D2.)

Another important point of consideration is the instructions that the panelists received during the standard setting process. For the NCEXTEND1 standard setting study, panelists were told not to consider guessing in their ratings. In other words, the cut scores recommended by standard setting panelists reflect the scores of

borderline examinees assuming the only reasons they scored correctly on items was that they were more likely than not to have known the answers. These two factors allow us to derive a reasonable system for adjusting for guessing.



Outcome	2 points	1 point	0 points
Probability	⅓	⅓	⅓

Expected score on an item (random guessing) = 2*(⅓) + 1*(⅓) + 0*(⅓) = 1

Figure D2. Visual explanation of the probabilistic outcomes for a given item if a student were to guess randomly

Consider a hypothetical cut score of 25. Under the idealized assumptions used in the standard setting study (where students only get items correct when they know the answers), there are three different ways a student could score a 25 (see Scenarios A-C, Table D1). A student could receive a score of 25 when not knowing the answer to either 0, 1, or 2 items. In the real world, a student would be unlikely to have a zero probability of getting the items correct that he/she did not know. Instead, the student would have a chance of getting an item correct due to guessing. Figure D1 shows that the average score a student would get on an item by guessing is 1. Assuming that each of the response patterns shown in Table D1 are equally likely, the average number of items that a student scoring 25 (based on knowledge alone) will not know the answer to, and thus would likely guess on is 1 (i.e., 0+1+2 divided by 3). Therefore, having established that the expected score on an item where guessing occurs is one, a reasonable adjustment-for-guessing for a cut score of 25 would be to increase the cut score by 1 point (i.e., an average of one item guessed on with the student receiving an average score of one) for an adjusted cut score of 26.

Table D1. Response patterns in which a student would score 25 points

Scenario	Number of items by Score Point		
	2	**1**	**0**
A	10	5	0
B	11	3	1
C	12	1	2

Adjustments for cut scores at other points can be obtained in the same manner. For example, consider a panel-recommended cut score of 16. There are eight different ways a student could achieve a score of 16 (see scenarios A-H, Table D2). Again, assuming that each of the response patterns are equally likely, the average number of items that a student scoring 16 (based on knowledge alone) will not know the answer to, and thus would likely guess on is 4 (i.e., 0+1+2+3+4+5+6+7 divided by 8, rounded to the nearest whole number). Therefore a reasonable adjustment-for-guessing with a cut score of 16 would be to increase the cut score by 4 points (i.e., an average of 4 item guessed on with the student receiving an average score of one for each of these items) for an adjusted cut score of 20.

Table D2. Response patterns in which a student would score 16 points

Scenario	Number of items by Score Point		
	2	**1**	**0**
A	1	14	0
B	2	12	1
C	3	10	2
D	4	8	3
E	5	6	4
F	6	4	5
G	7	2	6
H	8	0	7

This adjustment can be calculated in a similar manner for any score on the 0-30 range of the Extend1 assessments. Adjustments for lower scores will be larger than those for higher scores, using this methodology. See Table D3 for the full list of guessing adjustments.

Table D3. Guessing Adjustment by Score Point

Recommended Cut Score	Guessing Adjusted Cut Score
0	15
1	15
2	16
3	16
4	16
5	16
6	17
7	17
8	17
9	17
10	18
11	18
12	18
13	18
14	19
15	19
16	20
17	20
18	21
19	22
20	23
21	23
22	24
23	25
24	26
25	26
26	27
27	28
28	29
29	29
30	30

Appendix E: Evaluation Comments

Elementary Panel

- Larger paper to write ALDs during group time. Better training on the differences between a barely level student and a solid level student. Some participants don't work with the target population and don't have a clear understanding of their abilities- wide range between lowest and highest functioning students. These panelists often set level expectations too high. Performance information doesn't mean the students actually knew the answers. One has to factor in lucky guesses. Language of the questions has a huge effect on student performance. Vocabulary comprehension. There should be at least one other level of the Extend I. One for non-readers and students with lower cognitive functioning. This population has too wide of a range of ability. The stimulus cards are too big and often fall outside the students' range of vision. They don't often understand to look at all the choices unless prompted verbally and or gesturally. Overall it was a great experience. Wonderful staff- friendly and helpful and fun :)
- While panelist variation/diversity is important, there were a couple of members with little experience with the NCEXTEND I test and or the Extend I population of students. This process may or may not have been more successful with a more focused group with similar experiences. Additionally, a group with relevant and recent experience teaching the extended content standards may have produced more meaningful conversation and a move specific cut score based on SOUND knowledge of curriculum. DPI and Alpine were great in both facilitating and mediating opinions.
- The sessions were very engaging and helpful. Everything stayed on schedule and our professional opinions were valued and respected! Laura and Torrey did an amazing job of keeping us all on task and focused. In the elementary room I'm concerned some people did not have enough experience with the Extend I or teaching the 1% population so their info became outliers and their attitudes were not as positive. Laura was AWESOME at dealing with all of the crazy emotions and uproars!
- First, Laura did an excellent job sticking to the agenda/focus for the session! Torrey was a great help and everyone from Alpine/NCDPI was fantastic! The elem. group had some strong personalities and opinions which often made for a tense environment. Laura handled it beautifully and diplomatically. I think this experience has been enlightening and beneficial! I am concerned that there are some panel members with LITTLE to NO EXPERIENCE in Ext 1 who were in the elem. group (I cannot speak for the other groups)- but it concerns me that some specific people with a lack of experience or knowledge often seemed to be the outliers- or temperamental to the group! Thank you for the opportunity to participate in this process. I am happy to see the amount of time and information that is taken into consideration when making administrative decisions about high stakes testing.
- Teachers, for the most part, are used to working on their own time to get a task completed. If they know the task ahead of time the ALD process might be smoother and not require 2nd day revisions.

- This was an excellent experience. I'm glad there is a median, minimum and maximum score. Although I think some expectations are lofty. I feel the feedback given and ability to adjust will make scoring fair. Thank you for the opportunity.
- If possible, it would be helpful to have the training dates and schedule earlier than we did. I personally had many things to schedule all around each day of the trainings. All orientation and presentation pieces were very clear and well done. Thank you! Maybe show an example of a test from the get go for the few who have never administered. We even had a discrepancy in our room over the administration rules.
- I feel that panelists for the Extend I standard setting need to have more of a background working with this very specific population. Within my group, regular education teachers seemed to have some unrealistic expectations for this population.
- Wanted to know how much our input and work the last 3 days impacts possible changes in the Extend I. Felt it was a good process. One thought though- maybe it would be better to only have teachers who have taught this population. Reg ed teachers and spec ed seemed to have many differences and discrepancies.
- Screening of panelists needed to be based on the topic and with a minimum of so many years. Several panelists had not taught elementary level in many years which has changed with extended common core since then. Give more feedback on the way we could potential make changes to the Extend 1 test.
- Laura does a great job!
- I think in any group situation it is helpful to develop group norms before beginning- everyone seemed to work well but in this situation you want everyone to feel comfortable speaking up. In the room I was in there was some eye rolling and whispering when someone disagreed or challenged what was being stated.
- Everyone from Alpine testing and DPI were wonderful. Laura did an excellent job facilitating our group. I was surprised and a little concerned at the lack of background knowledge of the Extend 1 a handful of the panel had. At times, during creating ALDs these participants were usually ones with many questions and concerns. It concerns me when an elementary teacher says she does not know what I:I correspondence means. We had an ESL teacher, HS teacher, 1st grade and middle grades teacher. Got clarification from Torrey :) Also awesome!

Middle School Panel

- I enjoyed this thoroughly. Probably one of the MOST informative and productive workshops I have been involved with.
- What a powerful experience it has given me an understanding of how the tests and scores come about. Had a great group and worked very hard but the group was wonderful. I am pleased to chosen and have learned a lot from it.
- In the future it would be helpful to provide panelist with larger lined sheets of paper to record draft/final ALDs :)
- Possibly giving preliminary training about how Extend I is administered for those who have never seen it and are unaware of the accommodation/modifications allowed. I believe more emphasis should also be placed on the type of students and their disability ranges. Extend I students generally are VERY far from

their same age/grade peers. I also found it very difficult to develop ALDs based on standards that are WAY WAY too high for students with significant cognitive disabilities. The test we reviewed did not always align with ALDs too.

- Timing was off-- a little too much discussions. Perhaps next time we could be given time restraints. For example, "You have about 15 minutes to…" or "Let's discuss this for 5 minutes". Wonderful! Sara is a great facilitator: she kept us on task with a professionalism rarely seen this day and age. I loved working in groups to determine the strands and expectations for our students. I greatly appreciated being treated as a professional in my field. I felt honored being chosen to work on this task at hand.
- I really enjoyed and learned A LOT about how this process is completed. As a teacher, you are unaware of how standards are set and feel that it is a stab in the dark. This will assist me in going back home and being able to explain how intense the process and development/evaluation is!
- I would like for the group discussions between round 1 and 2 to be timed or facilitated where we stayed on task at hand. Consistent instructions on how to use standards to create an ALD. Example: Give middle school an example of a high school ALD standard of a barely 3. Thanks for the opportunity :)
- ALV- processing information- would liked to know we were doing this component so I could review information.
- More time to develop ADLs. More ability to choose content area most comfortable to develop ADLs for. More profiles of student for teachers unfamiliar with wide range of students in Extend 1 settings.
- Sara was great to work with, as were all the Alpine group! Everything was first class and enabled us to accomplish our goal. Thank you!

High School Panel

- Perhaps next time more effort could be placed in targeting EC teachers or specific teachers who work with NCExtend1 students. As a regular ed teacher, I did not always feel that I was very useful. BUT that being said, I learned a lot from simply experiencing how the process works. Thanks for the opportunity.
- Aside from going over on Wednesday, I found this program organized and efficient. I was grateful for the opportunity to see how achievement levels are determined. The presenters were engaging and effective in interpreting complex statistical processes in layman's language.
- More orientation and training for those unfamiliar with the Ext 1 population. More recruitment of Ext 1 personnel to assist in decision making and group discussion.
- Might be helpful to send invite to EC directors for counties- would potentially get more EC teachers involved for testing review.
- The most impactful problem in our group was the difference in opinion about whether standards or the level of the extend 1 students should determine the ALDs. Either standards matter and merit their use as dominant factors in writing ALDs or not. I recognize that the Extend 1 students are relatively low level, but the argument that their level should partially dictate the rigor of the ALDs ignores what the ALDs should represent as curriculum and testing guides. Our facilitator did appropriately mediate the dispute, but that schism tainted our ability to determine valid cut off scores. This matters! If a group consistently low balls student expectations, the cut offs will not dictate a rigorous and meaningful test. Either you get a test and curriculum that is watered down and essentially meaningless, OR you can

develop a curriculum template useful to course development and testing efficacy. As teachers, we all advocate for our students, but evaluation needs to be clinical and even handed. When student first individuals get too much sway in this process, they skew it away from a reality-based paradigm and limit its effectiveness. I do not see a way to bridge that divide when one teacher wants a 30, but another wants an 18. I am sure that it all works out somehow, but I do not know where DPI lies in this debate. The fact that some of these testing low ballers have been to other DPI standards setting meetings frankly terrifies me. I felt that, as a curriculum specialist, I might have been in the wrong place. In my mind, we needed a clear indication about which perspective was more important. Did we want a test curriculum all students can pass, or a rigorous approach to education? Either standards matter or they don't. Our moderator was GREAT and PROFESSIONAL and I appreciate his efforts. Very prepared and willing to work with us. Thank you!

- Lack of knowledge of the population of NCExtend I students by several participants hindered the progress or accuracy of the levels. There was a lot of confusion about whether level 4 should extend beyond the given standards. Some groups used the exact standards as level3 then extended level 4 beyond the id standards. Needed more time to go over the ALDs. Clarification on expectations for levels to standards. Need more time to vertical align with middle school. Provide examples of NC Extend I students so that participants are clear about population. Levels should consider standards and the population??? Felt a little rushed to finish both days even though plenty of time for Angoff-ALD confusion. Overall very informative, just needed more time and clarification. Thanks for this great opportunity.
- Felt high school 2nd day was too much, felt rushed and unable to really change ALDs when we disagreed completely with one of the groups. Felt some people that had no knowledge of this population hindered the process. Rating 4 hs tests 2 times in just a few hours was too much, more time needed when everyone was either upset or exhausted.
- It would be helpful to describe the various disabilities that the population of children who take the Extend I have and how the disability may affect their performance. Many of the participants in the standard setting workshop did not understand the students disabilities and test performance. The presenter did an excellent job of keeping the group under control and focused. The high school should have worked an additional day. We were rushed to complete our tasks.
- I feel there needs to be more ExI teachers involved. The ideal situation would be for reg ed content area teachers and ExI classroom teacher to be on this panel. There was also a communication breakdown between DPI and Alpine. There seemed to be a question as far as where the standards were targeted.
- Not enough people with true understanding of the Extend I population (one group had zero representation). Discrepancies between what standards mean in comparison to levels (some of the groups felt standards were barely 3 while other groups felt they were solid 4s). Cut scores needed to be based on barely levels but there was no discussion of what a barely was- facilitator stated barely 3 was a 3 when in reality they were a 2+. Did not discuss any changes to ALDs on day 2. No consensus to the charges group made never saw them until score rating. Rating scores showed some people felt level 3 and 4 were perfect score. Was told by 1 group person that it was the teachers fault if the kids didn't know at least what the standards state. Break process into 3 steps: ALLs solid level, ALDs barely levels, ratings based on ALDs barely levels. Need more time to discuss all questions rather than only looking at 1 easy and 1 hard. Biology and life science levels much higher expectations than ELA and algebra.

Needed another 1/2 day. Every round 1 have a maximum score of 30. 3 rounds of scoring with discussion. Cannot state how confident about scores without seeing final scores.