

The North Carolina Department of Public Instruction
Edition 4
Mathematics Grades 3 – 8 and NC Math 1
NCEXTEND1 Alternate Assessments
Technical Report
2018–2019



PUBLIC SCHOOLS OF NORTH CAROLINA
Division of Accountability Services | North Carolina Testing Program
Department of Public Instruction | State Board of Education

Prepared by:
North Carolina Department of Public Instruction
Accountability Division
December 2020
Tammy Howard, Ph.D., Director of Accountability
Kinge Mbella, Ph.D., Lead Psychometrician
Thakur Karkee, Ph. D., Psychometrician
Maxey-Moore, Section Chief, Test Development

In compliance with federal law, the NC Department of Public Instruction administers all state-operated educational programs, employment activities, and admissions without discrimination because of race, religion, national or ethnic origin, color, age, military service, disability, or gender, except where exemption is appropriate and allowed by law. Inquiries or complaints regarding discrimination issues should be directed to:

Ronald Paxton, Director of Human Resources
6301 Mail Service Center
Raleigh, NC 27699-6301
Telephone: (984) 236-2220

Table of Content

CHAPTER 1 INTRODUCTION	1
1. 1 Purpose and Background of the North Carolina State Testing Program	1
1. 2 NCEXTEND1 Content Standards Review, Revision and Implementation Processes	3
1. 3 Overview of the North Carolina Alternate Assessment Program	7
1. 4 Overview of the Technical Report	7
1. 5 Glossary of Abbreviations	9
CHAPTER 2 TEST DESIGN, ITEM DEVELOPMENT, AND FIELD-TEST PLAN	11
2.1 Test Specifications	11
2.1.1 Content Blueprint	11
2.1.2 Cognitive Complexity	13
2.1.3 Item Format	14
2.1.4 Calculator Use	15
2.2 Item Writer and Reviewer Training	15
2.3 Item Development Process	16
2.4 System Tryout Study	18
2.5 Mode of Test Administration	18
2.6 <i>Edition 4</i> NCEXTEND1 Mathematics Alternate Assessment Field-Test Design for Grades 3–8 and NC Math 1	19
CHAPTER 3 ITEM ANALYSIS	21
3.1 Statistical Item Flagging Criteria	21
3.2 CTT Based Item Analysis	22
3.3.1 Field-Test Sample Characteristics	25
3.3.2 Raw Scores and Timing Data	26
3.3 Fairness Review	27
CHAPTER 4 OPERATIONAL FORM ASSEMBLY, ANALYSIS, AND REVIEW	29
4.1 Form Assembly and Statistical Targets of New Forms	29
4.2 Form Review	30
4.2.1 Content Reviews	30
4.2.2 Production Reviews	31
4.2.3 Bias and Sensitivity Reviews	32
4.3 Summary of Final Operational Forms and Field-Test Design	33
4.3.1 <i>Edition 4</i> NCEXTEND1 Mathematics Test Structures and DOKs	34
4.3.2 Field-Test Plan	34
CHAPTER 5 TEST ADMINISTRATION	36
5.4.1 Protocols for Assessors	40

5.4.2	Protocol for Handling Alternate Assessments	40
5.5.1	Testing Windows.....	41
5.5.2	Modes of Test Administration.....	41
CHAPTER 6 SCORING AND SCALE DEVELOPMENT.....		51
6.1	Scoring and Scale Scores.....	51
6.2	Scaling	52
6.3	Automated Decentralized Scoring	53
6.4	Score Certification.....	53
CHAPTER 7 STANDARD SETTING		55
7.1	Standard Setting Activities	55
7.1.1	Panelists’ Backgrounds.....	56
7.1.2	Opening Session and Introductions	57
7.1.3	Achievement Level Descriptors	58
7.1.4	Method and Procedure.....	59
7.1.5	Across-Grade Articulation and Final ALD Cuts.....	59
7.2	Evaluation of the Standard Setting Workshop.....	60
7.2.1	Participants’ Evaluation.....	61
7.2.2	External Evaluation	61
CHAPTER 8 TEST RESULTS AND REPORTS		63
8.1	NCEXTEND1 Scale Score Distribution.....	63
8.1.1	Scale Scores by Disability Subgroups	67
8.1.2	Scale Scores by Gender	69
8.1.3	Scale Score by Major Ethnic Groups.....	71
8.1.4	Achievement Levels Distributions	72
8.2	Score Reports	74
8.3	Confidentiality of Student Information	75
8.3.1	Confidentiality of Personal Information.....	76
8.3.2	Confidentiality of Test Data.....	76
CHAPTER 9 VALIDITY EVIDENCES.....		78
9.1	Reliabilities of the NCEXTEND1 Assessments	78
9.2	Conditional Standard Errors at Scale Score Cuts.....	79
9.3	Classification Consistency	80
9.4	Unidimensionality of NCEXTEND1 Assessments.....	82
9.4.1	Eigenvalues and Variance.....	82
9.5	Alignment Study	87
9.6	Fairness and Accessibility	87
9.6.1	Accessibility in Universal Design	87
9.6.2	Fairness in Access	88

9.6.3 Fairness in Administration89

Glossaries 90

References 92

Table of Tables

Table 1. 1 NC Math 1 Extended Content Standards Review, Revision, and Implementation
Timeline.....4

Table 1. 2 Grades K – 8 NCEXTEND1 Mathematics Standards Review, Revision and
Implementation Timeline.....5

Table 1. 2 NCDPI Accountability and Testing Timeline Highlights, Mathematics
NCEXTEND16

Table 2. 1 NC Math 1 NCEXTEND1 Test Blueprint (%).....12

Table 2. 2 NCEXTEND1 Mathematics Test Blueprint (%), Elementary and Middle School ...13

Table 2. 3 Proposed Target DOKs Across Grades.....14

Table 2. 4 Grades 3–8 and NC Math 1 NCEXTEND1 Mathematics Standalone
Field Test Plan, 2017–1819

Table 3. 1 CTT Item Flagging Criteria.....23

Table 3. 2 CTT Descriptive Summary of Field-Test Item Pool, Spring 201824

Table 3. 3 Demographic Distribution (%) of 2017–18 Mathematics Field-Test Population25

Table 3. 4 Distribution (%) of NCEXTEND1 2017–18 Mathematics Field-Test Student
Population by Disability Categories26

Table 3. 5 Grades 3–8 and NC Math 1 FT Descriptive Statistics of Raw Score
and Timing Data27

Table 4. 1 Field-Test CTT Summary Statistics, Spring 2018.....30

Table 4. 2 Demographic Information for Fairness Review Panels, Spring 201833

Table 4. 3 Test Structures, Grades 3–8, and NC Math 134

Table 4. 4 Grades 3–8 and NC Math 1 Field-Test Design35

Table 5. 1 Recorded Test Duration (Minutes) for Grades 3–8 and NC Math 1 Operational
Forms, 2018–1944

Table 5. 2 Approved accommodations for the NCEXTEND1 Alternate Assessments45

Table 5. 3 Participation of Students (%) by Assessment – General and NCEXTEND146

Table 6. 1 Average CTT Statistics, 2018–19 Operational Forms.....52

Table 7. 1 Panelist Gender and Ethnicity 56

Table 7. 2 Panelist Experience as Educators (%) 56

Table 7. 3 Panelist Professional Background: Three–Grade Panels 57

Table 7. 4 Panelist Professional Background: Single–Grade Panels 57

Table 7. 5 Policy ALDs for NCEXTEND1 Alternate Mathematics Assessment 58

Table 7. 6 Final NCEXTEND1 Recommended Cuts and Proficiency Distributions 60

Table 7. 7 2018-19 Population NCEXTEND1 Raw Score Ranges Across Achievement Levels 60

Table 7. 8	NCEXTEND1 Standard Setting Workshop Evaluation Results	61
Table 8. 1	Grades 3–5 NCEXTEND1 Mathematics Scale Score by Accommodation Subgroups, Spring 2019.....	68
Table 8. 2	Grades 6–8 and NC Math 1 NCEXTEND1 Mathematics Scale Score by Accommodation Subgroups, Spring 2019	69
Table 8. 3	Grades 3–5 NCEXTEND1 Mathematics Scale Score by Gender, Spring 2019	70
Table 8. 4	Grades 6–8 and NC Math 1 NCEXTEND1 Mathematics Scale Score by Gender, Spring 2019	70
Table 8. 5	Grades 3–4 NCEXTEND1 Mathematics Scale Score by Ethnicity, Spring 2019	71
Table 8. 6	Grades 6–8 and NC Math 1 NCEXTEND1 Mathematics Scale Score by Ethnicity, Spring 2019	72
Table 8. 7	Reports by Audience.....	75
Table 9. 1	Overall and Subgroup Reliabilities, NCEXTEND1 Mathematics	79
Table 9. 2	NCEXTEND1 Conditional Standard Errors at Achievement Level Cuts	80
Table 9. 3	Classification Accuracy and Consistency Results, NCEXTEND1 Mathematics	81
Table 9. 4	Eigenvalues (Eigen) and Variance (%) Accounted for by the Components, Grades 3–8 and NC Math 1.....	86

Table of Figures

Figure 8. 1	Grade 3 NCEXTEND1 Mathematics Scale Score Distribution, Spring 2019	64
Figure 8. 2	Grade 4 NCEXTEND1 Mathematics Scale Score Distribution, Spring 2019	64
Figure 8. 3	Grade 5 NCEXTEND1 Mathematics Scale Score Distribution, Spring 2019	65
Figure 8. 4	Grade 6 NCEXTEND1 Mathematics Scale Score Distribution, Spring 2019	65
Figure 8. 5	Grade 7 NCEXTEND1 Mathematics Scale Score Distribution, Spring 2019	66
Figure 8. 6	Grade 8 NCEXTEND1 Mathematics Scale Score Distribution, Spring 2019	66
Figure 8. 7	NC Math 1 NCEXTEND1 Mathematics Scale Score Distribution, Spring 2019	67
Figure 8. 8	State Level NCEXTEND1 Achievement Level Classifications by Grade, Spring 2019	73
Figure 8. 9	Individual Student Report (ISR).....	74
Figure 9. 1	Grade 3 NCEXTEND1 Mathematics Scree Plots of 2018–19 Operational Forms...83	
Figure 9. 2	Grade 4 NCEXTEND1 Mathematics Scree Plots of 2018–19 Operational Forms...83	
Figure 9. 3	Grade 5 NCEXTEND1 Mathematics Scree Plots of 2018–19 Operational Forms...84	
Figure 9. 4	Grade 6 NCEXTEND1 Mathematics Scree Plots of 2018–19 Operational Forms...84	
Figure 9. 5	Grade 7 NCEXTEND1 Mathematics Scree Plots of 2018–19 Operational Forms...85	
Figure 9. 6	Grade 8 NCEXTEND1 Mathematics Scree Plots of 2018–19 Operational Forms...85	
Figure 9. 7	NC Math 1 NCEXTEND1 Mathematics Scree Plots of 2018–19 Operational Forms	86

CHAPTER 1 INTRODUCTION

Every Student Succeeds Act (ESSA) requires participation of students with the most significant cognitive disabilities in the Alternate Assessment aligned to alternate academic achievement standards (AA-AAAS) without exceeding 1.0 percent of the total number of students in the state who are assessed in that subject. Students with significant cognitive disabilities have cognitive and adaptive behavior functioning deficits that may prevent them from attaining grade level achievement standards, even with substantial program modifications and accommodations. They may require extensive individualized instruction across multiple settings to access and make progress in the learning environment.

The North Carolina State Board of Education (NCSBE) policy ACCT-021 states that all eligible students enrolled in a North Carolina school at grades 3–8 and NC Math 1 high school courses shall participate in the state assessment program adopted by NCSBE. All students with disabilities who are enrolled in a school, including those with Individual Education Programs (IEPs) and those identified under Section 504 of the Rehabilitation Act of 1973, shall participate in the state assessment program using one of the following assessments as appropriate and as determined by the student’s IEP or Section 504 Plan:

- The standard test administration with or without accommodations, or
- An Alternate Assessment with or without accommodations, if eligible.

The grades 3–8 Alternate Assessments based on the North Carolina Extended Content Standards (ECS) for Mathematics administered by the North Carolina Department of Public Instruction (NCDPI) measure students’ proficiency on the ECS. The assessment results are used for school and district accountability under the accountability model and for federal reporting purposes.

The intent of this technical report is to provide comprehensive and detailed evidence in support of the validity and reliability of the North Carolina NCEXTEND1 Alternate Assessments. The first part of this report presents a brief overview of the revision and eventual adoption of the new NCEXTEND1 mathematics content standards which are used to justify the development of new assessments. The remaining sections describe a brief history of the NCEXTEND1 Alternate Assessments followed by documentation of item development and review, field test and analysis, and form development and review. The report concludes with summaries of standard setting workshop used to set achievement levels for reporting and interpreting student results, and validity evidence for the *Edition 4* NCEXTEND1 mathematics summative Alternate Assessments.

1.1 Purpose and Background of the North Carolina State Testing Program

The General Assembly GCS 115C-174.10T specified the purpose of the North Carolina Annual Testing Program (NCATP) as:

“(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 the above purposes as a guide, the NCSBE developed the School-Based Management and Accountability Program to improve student performance in the early 1990s. The current vision of the NCSBE is “*Every public school student will be empowered to accept academic challenges, prepared to pursue their chosen path after graduating high school, and encouraged to become lifelong learners with the capacity to engage in a globally-collaborative society.*” The current mission of the NCSBE is to use its constitutional authority to guard and maintain the right of sound, basic education for every child in North Carolina Public Schools including students with the most significant cognitive disabilities. The NCSBE’s three main goals are to:

- *Eliminate opportunity gaps by 2025,*
- *Improve school and district performance by 2025, and*
- *Increase educator preparedness to meet the needs of every student by 2025.*

Starting from the early 1990s, North Carolina has continually sought innovation in the design, development, and ways to use state assessments to increase academic expectations, so students are prepared for success after high school. This is evident in the NCSBE stated goals and policy of continuous academic content standards evaluation and review. The NCSBE mandates that the NCDPI review content standards every five to seven years after they were first adopted. This also implies that state assessments are also reviewed and redesigned to ensure they are up to date with current measurement practices and aligned to academic expectations of current NCEXTEND1 North Carolina Extended Content Standards.

History of mathematics NCEXTEND1 Alternate Assessments dates back to 2005–06 where students with the most significant cognitive disabilities on grade-level ECS were tested in mathematics at grades 3–8 and NC Math 1. 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 Alternate Assessments, which was first administered in 2007–2008 administration cycle. *Edition 3* of the NCEXTEND1 assessments was adopted in February 2011 and implemented as a field test in the spring of 2012 by aligning all items to the NCECS for mathematics grades 3–8, and Algebra I A and B. The Algebra I A and B course approved by the NCSBE has been since renamed to NC Math 1, although the standards for the course remain the same. Based on test administration observations, data analysis, and teacher feedback, the scoring model included scaffolding test item administration where students were provided up to two trials per item to select the correct response.

The North Carolina mathematics content standards were revised in 2016–17. In order to maintain a strong content alignment with the North Carolina *Standard Course of Study* (NCSCoS), the

Edition 3 NCEXTEND1 Alternate Assessments were also revised, redesigned and operationally administered as *Edition 4* NCEXTEND1 Alternate Assessments in 2018–19. The redesigned assessments contained two parts with more items in the test.

The NCEXTEND1 Alternate Assessments are teacher-facilitated online assessments that are administered individually to each student. For grades 3–8 and NC Math 1 NCEXTEND1 Alternate Assessments, the questions are presented online in two (2) sets. There is no formalized break between the sets. Students have two (2) trials each for the first and second set of test questions. If the student misses the answer in the first trial, the incorrect answer is removed from the choices for the second trial. The test will terminate at the end of set one for students who do not answer enough questions correctly. Students who answer enough questions correctly in set one will continue to set two.

This technical report documents all steps and processes that were implemented in the development, administration, scoring, and reporting of results for *Edition 4* of NCEXTEND1 alternate mathematics assessments. The purpose of this report is to demonstrate the NCDPI’s continuous commitment to the highest standards and technical quality of its NCEXTEND1 Alternate Assessments.

1.2 NCEXTEND1 Content Standards Review, Revision and Implementation Processes

The *Edition 4* NCEXTEND1 Alternate Assessments followed the same standards review, revisions, and implementations process as the general assessments. The Exceptional Children Division (EC), in collaboration with K-12 Standards, Curriculum and Instruction Division (SC&I), developed and implemented a plan of action and timeline in 2016 to review and revise the Mathematics ECS.

Tables 1.1 and *1.2* outline timelines and brief descriptions of actions that were implemented by the NCDPI during the review, revision, and implementation of the new Mathematics ECSs from 2017 through 2019. *Table 1.1* shows timelines for the high school NC Math 1 Mathematics ECS. *Table 1.2* shows the timeline for grades 3–8 Mathematics ECS. These timelines show how the four principles - feedback based, research informed, improvement oriented, and process driven - outlined by the NCSBE were operationalized and implemented into actionable steps during the review, revision, and implementation of the new mathematics ECS.

Table 1.1 NCEXTEND1 NC Math 1 Extended Content Standards Review, Revision, and Implementation Timeline

Date	Actions	Descriptions
April 21, 2017	Alternate Assessment redesign focus group Grades 3–8, NC Math 1: English Language Arts (ELA)/Reading, Math, and Science	The Focus Group reviewed assessable Extended Content Standards, manipulatives, and content specific topics.
May 19, 2017	NC Math 1 Alternate Assessment test specifications meeting	Brief review of the current NCEXTEND1 design; identification and prioritization of assessable standards; and recommendation of administration time, number of test questions, and calculator use.
July 3, 2017	NC Math 1 Alternate Assessment-content connectors	The participants identified core academic content by standard and recommended content connectors. They also identified assessable standards based on content connectors as well as standards that may be integrated for assessment and recommended cognitive complexity (Depth of Knowledge) by standard and item types.
July 10–11, 2017	NC Math 1 Alternate Assessment item writing meeting	Participants discussed: <ul style="list-style-type: none"> - NC Testing Program - Alternate Assessment Basics - Eligibility Criteria and Student Characteristics - Universal Design - Diversity, Fairness, and Sensitivity Guidelines - Alternate Assessment Test Design - Item Development Guidelines and Examples - NC Math 1 Extended Content Standards - Review template - Practice Item Writing for Group - Review items from Groups
October 20, 2017	Extended Content Standards item review training	Participants discussed: <ul style="list-style-type: none"> - Alternate Assessment basics - Student characteristics - Alternate Assessment test design - Review of extended content standards and depth of knowledge (DOK)

Date	Actions	Descriptions
		- Item development and review basics practice item review.
2017–18	Field-Test	The NCDPI conducted a standalone field-test.
2018–19	Implementation	The NCDPI developed NCEXTEND1 mathematics alternate operational assessments (<i>Edition 4</i>) forms for administration.

Table 1.2 *Grades 3 – 8 NCEXTEND1 Mathematics Standards Review, Revision and Implementation Timeline*

Date	Actions	Descriptions
April 21, 2017	Alternate Assessment redesign focus group - Grades 3–8, NC Math 1: Reading, Mathematics, and Science	The Focus Group reviewed Extended Content Standards, manipulatives, and content specific topics.
September 6–7, 2017	Extended Content Standards for grades 3–8 mathematics approved.	In fulfillment of policy <i>SCOS-012</i> , the grades 3–8 mathematics ECS, were approved by the September 2017 NCSBE meeting.
September 16, 2017	Grades 3–8 mathematics Alternate Assessments specifications meeting	Participants reviewed current mathematics grades 3–8 NCEXTEND1 design; identified and prioritized assessable standards, and recommended administration time, number of test questions, and calculator use.
September 23, 2017	Grades 6–8 Mathematics Alternate Assessment item writing training	Participants discussed NC Testing Program; Alternate Assessment basics; eligibility criteria and student characteristics; universal design, diversity, fairness, and sensitivity guidelines; Alternate Assessment test design; item development guidelines and examples, grades 6–8 mathematics Extended Content Standards; practice item writing; and online test development system (TDS).
October 7, 2017	Grades 3–5 mathematics Alternate Assessment item writing training	Participants discussed NC Testing Program; Alternate Assessment basics; eligibility criteria and student characteristics; universal design, diversity, fairness, and sensitivity guidelines; Alternate Assessment test design; item development guidelines and

Date	Actions	Descriptions
		examples, grades 6–8 mathematics Extended Content Standards; practice item writing; and online test development system (TDS).
October 20, 2017	Extended Content Standards item review training	Participants discussed Alternate Assessment basics; student characteristics; Alternate Assessment test design, review of Extended Content Standards and depth of knowledge (DOK); item development and review of basics practice item review.
2017–18	Field-Test	The NCDPI conducted a standalone field test.
2018–19	Implementation	Develop NCEXTEND1 mathematics operational test (<i>Edition 4</i>) forms for administration.

The attributes described above are a part of validity evidence to show that North Carolina mathematics ECS are research based and have adequate rigor and expectation to prepare students with significant cognitive disabilities for post-secondary outcomes. To maintain content and construct validity evidence of NCEXTEND1 Alternate Assessment score uses and interpretation, North Carolina redesigned and administered new assessments that are aligned to the new adopted mathematics Extended Content Standards. *Table 1.3* shows an overview of the timeline beginning with adoption of new Extended Content Standards to development and reporting of scores aligned to these new mathematics content standards.

Table 1.3 NCDPI Accountability and Testing Timeline Highlights, Mathematics NCEXTEND1

Year	Action
June 2016	The NCSBE adopted the revised standards for EOC Mathematics.
August 2016	The revised standards for EOC mathematics were implemented.
January 2017	The NCSBE approved the new Extended Content Standards for NC Math 1.
September 2017	The NCSBE approved the new Extended Content Standards for grades 3–8.
2017–18	A stand-alone field-test was administered for both grades 3-8 and NC Math 1 Alternate Assessments.
2018–19	Operational forms (<i>Edition 4</i>) developed and administered for the first time.

1.3 Overview of the North Carolina Alternate Assessment Program

The NCDPI designs, develops, and administers customized high-quality statewide NCEXTEND1 mathematics Alternate Assessments in grades 3–8 and NC Math 1 that are aligned to the newly revised North Carolina ECS adopted by the NCSBE in 2017. The NCEXTEND1 Alternate Assessment scores provide valid and reliable information intended to serve two general purposes: measures students’ performance on the mathematics ECS and progress as it relates to their proficiency towards grade-level content standards and serves as a quantitative indicator for use in federal and statewide accountability models.

- Measure students’ performance and progress: The NCEXTEND1 Alternate Assessments measure the full breadth and depth of grade-level students’ performance on the North Carolina ECS for Mathematics. Student performance on the NCEXTEND1 Alternate Assessments is reported using scale scores grouped into one of three achievement levels (Not Proficient, Level 3, and Level 4). Level 3 is considered grade level proficiency and Level 4 is considered on track for competitive employment and postsecondary education.
- Federal and State Accountability and Reporting: The NCEXTEND1 Alternate Assessments are used as part of the assessment indicators in the school accountability model and included in assessment reporting as required by federal and state law. The accountability model is designed to identify schools in need of support. Performance reporting identifies the percentage of students in the school who score at each of the Achievement Levels. When possible, the performance reporting is disaggregated by regular and Alternate Assessments to provide information on the performance of students with significant cognitive disabilities.

The North Carolina *Testing Code of Ethics (Appendix 1-A)* cautions educators to use the NCEXTEND1 Alternate Assessment scores and reports only for these intended uses as approved by the NCSBE and for which NCDPI has provided validity evidence to support these intended uses. It also reiterates that test scores are only one of many indicators of student achievement. The use of the NCEXTEND1 Alternate Assessment scores for purposes other than those intended by NCDPI must be supported by evidence of validity, reliability, precision, and fairness.

1.4 Overview of the Technical Report

Validity is a unifying and core concept in test development processes. Therefore, validity evidence of the NCEXTEND1 Alternate Assessments are documented throughout this report.

Chapter 1 provides a brief history of testing in North Carolina; the standards review, revision and implementation process; and overview of the North Carolina statewide assessment program.

Chapter 2 documents an overview of NCATP test design, item development process and field-test plans. The test design involved test specifications meetings to specify test blueprints, test complexities, calculator use, item format, and mode of test administration. The item development process involved item writer training, item writing, and reviews. Final sections describe field-test plans to replenish the item pool for future test development.

Chapter 3 describes the field-test item analysis plans using Classical and Item Response theory as well as differential item functioning analysis. NCDPI has set internal criteria for filtering out items with less-than-optimal quality. Final sections describe summary of item analysis and separate and concurrent calibration of item responses administered in paper and computer modes.

Chapter 4 starts with form assembly process using *Edition 3* classical test theory based preliminary statistical targets. In subsequent sections, descriptions of 24-step operational form assembly and review processes are documented. One form is developed for each of the grades, based on the field-test statistics that are closely aligned to each other, content-wise in terms of blueprint and complexity, and statistically in terms of proportion correct (p-value) and biserial correlations. The forms are then reviewed by content experts at the NCDPI and North Carolina State University Technical Outreach for Public Schools (TOPS) for overall quality, outside content experts for independent evaluation, and by psychometrician for key balance as well as verifying content and statistical specifications. The chapter also documents the structure of the base forms in terms of item types and text complexities, and descriptive classical statistics based on the field-test data.

Chapter 5 documents procedures put in place by the NCDPI to assure the administration of NCEXTEND1 assessments are standardized, fair and secured for all students across the state. The chapter also describes test administration training to assessors, test security and accommodation procedures implemented to ensure all students with disabilities and English Learners (ELs) are able to take NCEXTEND1 assessments. The chapter concludes with description of student participation and processes used for identifying test irregularities and misadministration.

Chapter 6 describes processes used for scoring and scale development procedures adopted to create final reportable scale scores. The chapter begins with documenting final CTT statistics and scaling. Final sections describe score certification process.

Chapter 7 presents a summary of the standard setting study that was conducted in July 2019 after the first operational administration of NCEXTEND1 tests. The NCDPI contracted with Data Recognition Corporation (DRC) to conduct a standard setting workshop to recommend cut scores and achievement levels for the newly developed mathematics NCEXTEND1 assessments. The chapter is a condensed version of the final report prepared by DRC describing the full workshop and final cut score recommendations. Final section documents validity of the standard setting in terms of participants’ evaluation of standard setting processes as well as evaluation of the process by external evaluators.

Chapter 8 presents summary student performance results for the 2018–19 NCEXTEND1 operational assessments. This chapter is organized into three main sections. The first section highlights descriptive summary results of scale scores and achievement levels for the NCEXTEND1 tests across major demographic variables. The second section presents sample reports and descriptions and stakeholders of the various standardized reports created by the NCDPI. The final section briefly describes confidentiality of student information.

Chapter 9 presents summary validity evidence collected in support of the interpretation of the NCEXTEND1 test scores. The first two sections in this chapter present validity evidence in support of internal structure of the NCEXTEND1 assessments. Evidence presented in these sections includes reliability, standard error estimates, classification consistency summary of reported achievement levels and exploratory Principal Component Analysis in support of the unidimensional analysis and interpretation of scores. The final section presents a summary of procedures used to ensure the NCEXTEND1 assessments are accessible and fair to all students.

1.5 Glossary of Abbreviations

Abbreviations	Full Form
ALD	Achievement Level Descriptor
ASRC	Academic Standards Review Commission
CBT	Computer-Based Test
CTT	Classical Test Theory
DIF	Differential Item Functioning
DOK	Depth of Knowledge
DRC	Data Recognition Corporation
EC	Exceptional Children
ECS	Extended Content Standards
EDS	Economically Disadvantaged Students
EL	English Learner
ELA	English Language Arts

Abbreviations	Full Form
EOC	End-of-Course
EOG	End-of-Grade
ESL	English as a Second Language
ESSA	Every Student Succeeds Act
FERPA	Family Educational Rights and Privacy Act
HOSS	Highest Obtainable Scale Score
IDEA	Individuals with Disabilities Education Improvement Act
IEP	Individualized Education Program
IRT	Item Response Theory
LEA	Local Education Agency
LOSS	Lowest Obtainable Scale Score
MC	Multiple Choice
NC	North Carolina
NCDPI	North Carolina Department of Public Instruction
NCEXTEND1	North Carolina Alternate Assessment
NCSBE	North Carolina State Board of Education
<i>NCSCOS</i>	North Carolina Standard Course of Study
NCATP	North Carolina Annual Testing Program
NCSU-TOPS	North Carolina State University-Technical Outreach for Public Schools
NCTAC	North Carolina Technical Advisory Committee
OTISS	Online Testing Irregularity Submission System
PBT	Paper-Based Test
PCA	Principle Component Analysis
PII	Personally Identifiable Information
NCDPI-SC&I	Standards, Curriculum, and Instruction
SE	Standard Error
SWD	Students with Disabilities
TDS	Test Development System
TMS	Test Measurement Specialist
VI	Visually Impaired

CHAPTER 2 TEST DESIGN, ITEM DEVELOPMENT, AND FIELD-TEST PLAN

This chapter documents steps implemented by the NCDPI during the development of *Edition 4* NCEXTEND1 mathematics assessments in adherence with Standard 4.0 (AERA, APA, & NCME, 2014) which 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). Specifically, this chapter describes the test specification processes – content blueprint, test format, item development, and review. The last section describes the item tryout plans used to field-test newly developed items for NCEXTEND1 *Edition 4* Alternate Assessments.

2.1 Test Specifications

The NCEXTEND1 grades 3-8 and NC Math 1 are standard-based Alternate Assessments that serve summative purposes. These Alternate Assessments were redesigned from focus group meetings and they were aligned with new mathematics content standards adopted in 2016–17 to ensure adequate validity evidence in support of standard-based interpretation of test scores. The second step in the development of the new assessments is guided by the overall test specifications which outline all essential content, cognitive, and psychometric specifications.

The NCDPI recruited North Carolina teachers and educators from across the state and conducted two on-site test specification workshops in 2017 for both the general and Alternate Assessments. Educators invited to these meetings represented North Carolina educators and teachers from across all geographic regions, demographic subgroups, and experiences. Participants also included special education and English Learner educators to ensure fairness and accessibility of Alternate Assessments for all North Carolina students. Full agendas, surveys, and complete demographic characteristics of workshop participants by grade span are tabulated in [Appendix 2–A](#). The main purposes of these test specification workshops were to specify content, cognitive, test format blueprints and psychometric specifications for *Edition 5* general assessments and *Edition 4* Alternate Assessments.

2.1.1 Content Blueprint

The two-day on-site test specification workshops facilitated by NCDPI test development staff were designed to get participants to recommend content blueprints for general *Edition 5* and NCEXTEND1 *Edition 4*. The workshops for NCEXTEND1 were held by grade spans: NC Math 1 and grades 3–8. During these workshops, participants were tasked to recommend content

domain blueprints for each grade. Workshops started with an overview presentation of the purposes of NCEXTEND1 Alternate Assessments followed by an overview of the new mathematics extended content standards. Participants were then separated into smaller work groups, and each group was assigned a group lead to facilitate discussions. The first major task for participants was to recommend content blueprint weights by domain. These recommendations were done in two rounds with large group discussions between rounds.

In Round 1, following group discussions of grade level content standards as they relate to the NCEXTEND1 Alternate Assessments, participants were directed to individually assign 0–10 ratings on a Google form with “0” indicating a particular standard cannot be assessed based on the proposed assessment design to “10” indicating a standard can be assessed and is of the highest importance. At the conclusion of Round 1, all ratings were aggregated and summarized to generate recommended domain content distribution weights.

The Round 1 recommendations from all participants were aggregated and presented to the larger group for open discussions. Group discussions were prioritized for standards with the highest ranges of ratings among participants. During these group discussions participants were given an opportunity to justify their ratings and share their rationale with the entire room. Following large group discussions, participants returned to their smaller groups for one final round of recommendations.

In Round 2, participants were encouraged to rely on information shared from the larger group discussions to determine if they wanted to revise any ratings. At the conclusion of Round 2 reviews, the updated recommended content weights were presented as their final grade-level content blueprint recommendations.

At the end of test specification workshops, the NCDPI team members from Test Development and subsequently Standards, Curriculum, and Instruction reviewed the recommended blueprints for the extended content standards to ensure adequate across-grades articulation. The final recommendations shown in *Tables 2.1 and Table 2.2* were then adopted as *Edition 4* mathematics content blueprints for NCEXTEND1 assessments.

Table 2.1 NCEXTEND1 NC Math 1 Alternate Assessment Test Blueprint (%)

Domain	NC Math 1
Number and Quantity and Algebra	38–42
Functions	28–32
Geometry	12–16
Statistics	17–21

Table 2.2 NCEXTEND1 Mathematics Test Blueprint (%), Elementary and Middle School

Domain	Grades		
Elementary	3	4	5
Operations and Algebraic Thinking	25–29	26–30	11–14
Number and Operations in Base Ten	22–26	19–23	42–45
Number and Operations - Fractions	14–18	14–18	11–15
Measurement and Data, Geometry	34–37		
Measurement and Data		14–18	15–19
Geometry		19–23	19–22
Middle School	6	7	8
Ratios and Proportional Relationships	15–19	11–15	
The Number System	26–30	19–22	15–19
Expressions and Equations	19–22	15–19	25–29
Functions			15–19
Geometry	15–19	30–33	18–22
Statistics and Probability	15–19	19–22	11–15

2.1.2 Cognitive Complexity

On Day 2 of the test specification workshop, participants were tasked to evaluate and recommend cognitive complexity expectation ranges for all assessable standards to guide item and test development. The NCDPI adopted the Norman Webb Depth of Knowledge (DOK) classification (Hess, 2013) as the basis for evaluating complexity for NCEXTEND1 Alternate Assessment items. A general definition for each of the four DOK levels is shown in [Appendix 2–B](#). The DOK levels offer a framework for content experts to differentiate learning expectations and outcomes by considering the level of thinking required by students to successfully engage with items aligned to specific content standard expectations. Prior to the test specification workshops, NCDPI Test Development and SC&I staff received training on Webb’s DOK classifications on April 2017 from Dr. Karen Hess. The Webb’s DOK levels guide used during training by Dr. Hess is shown in [Appendix 2–C](#).

At the test specification workshop, the NCDPI staff provided an overview training on Webb’s DOK to ensure participants had the necessary working knowledge needed for this activity. They then participated in two rounds of discussions and recommendations of DOK expectations. In Round 1, participants were separated into smaller working groups and their task was to set DOK range expectations by standards. Classification ratings from each group were recorded using Google forms and the final data from all groups were uploaded into a final table and

reviewed with the entire large group. The large group discussions were used to give participants an opportunity to review and justify their ratings and make any necessary changes.

The final recommended DOK classifications from Round 2 were then adopted as the expected cognitive complexity recommendations for assessed mathematics content standards. At the conclusion of the meeting, the NCDPI’s Test Development and SC&I reviewed these recommended classifications to ensure coherent alignment with grade level content standards expectations and summarized the data into DOK range specifications for the NCEXTEND1 Alternate Assessments. The final cognitive complexity specifications for *Edition 4* mathematics NCEXTEND1 Alternate Assessments are shown in *Table 2.3*.

Table 2.3 Proposed Target DOKs Across Grades

Grade	Number of Items	Category (%)	
		DOK 1	DOK 2
3	27	60–70	30–40
4	27	60–70	30–40
5	27	60–70	30–40
6	27	50–60	40–50
7	27	50–60	40–50
8	27	50–60	40–50
10	28	50–60	40–50

2.1.3 Item Format

The NCEXTEND1 grades 3–8 and NCEXTEND1 NC Math 1 Alternate Assessments are teacher-facilitated online assessments that are administered individually to each student. The item format is three-option selected response presented online in two (2) sets. There is no formalized break between the sets. Students have two (2) trials each for the first and second set of test questions. If the student answered correctly in the first trial, he/she scores 2-point. If the student misses, the incorrect answer is removed from the choices for the second trial. If the student answered the item correctly in the second trial, he/she scores 1-point. Otherwise, the student scores zero. The test will terminate at the end of set one for students who do not answer enough questions correctly. Students who answer enough questions correctly in set one will continue to set two.

Rationales for using three-option selected response items are driven by practical and policy considerations. Practically, scaffolding has been used in class for engaging and obtaining a level of understanding of the significant cognitive disability students in North Carolina. The NCDPI policies are directed towards ensuring state assessments have a minimum effect on instructional time and resources yet are still able to guarantee reliable score for valid uses.

2.1.4 Calculator Use

The minimum calculator requirement for the mathematics tests in grades 6–8 and NC Math 1 is any four-function calculator with memory key. Students may use calculators with more than the minimum requirements (e.g., fraction keys), provided that those additional features are not prohibited. However, school systems should be cautioned that the use of these features without prior training may confuse students and adversely affect their performance during the test administration.

Students are to use calculators that are similar to those used during classroom instruction within the current school year. Students who regularly use more than one calculator during classroom instructional activities may be permitted to use more than one calculator during the test administration. Students are not allowed to use calculators with wi-fi access or built-in communication devices.

2.2 Item Writer and Reviewer Training

The first step of item development is item writer and reviewer training. The main pool of item writers and reviewers for the NCEXTEND1 Alternate Assessments are classroom teachers from North Carolina. Teachers who want to serve as item writers or reviewers are required to successfully complete in-person or online training courses available through the NC Education website: (<https://center.ncsu.edu/ncpd/course/>). The training includes a general course on item writing guidelines, including lessons on sensitivity and bias concerns. These courses provide an overview of the test development process and the basic rules and structures of item formats used by the North Carolina Testing Program. Those interested in item writing and/or reviewing should complete an application for becoming an item writer or reviewer.

The design of these courses is generally linear, requiring the online participant to step through each resource (Web page, PDF, etc.) in a structured sequence. At the end of most topic areas, participants are required to take a short quiz before moving to the next topic area to demonstrate understanding of the presented material. All online quizzes may be taken as many times as needed in order to meet the requirements for moving forward in the course. Once participants have viewed a resource, they are able to return to it for reference at any time. The online item writer training courses can be accessed using the website after login access.

Item writer and reviewer training also incorporates the concept of universal design and comprehensible access to the content being measured. For more information regarding the item writer training and how to become an item writer or reviewer for the North Carolina student testing program, visit the [website](#).

2.3 Item Development Process

The item development process for the mathematics NCEXTEND1 *Edition 4* Alternate Assessments began after the NCSBE adopted the new NCECS for NC Math 1 in June 2016 and for grades 3–8 in June 2017. North Carolina test items are written and reviewed by trained North Carolina teachers who serve as item writers. Additionally, Test Measurement Specialist (TMS) in partnership with Content Specialists at North Carolina State University Technical Outreach for Public Schools (NCSU-TOPS) at North Carolina State University participated in the item development processes. Ultimately, the NCDPI’s TMSs serve as the final staff reviewer for all NCEXTEND1 Alternate Assessment items. Educators with classroom and grade level Extended Content Standards experience across the state are recruited, trained, and awarded contracts to write NCEXTEND1 Alternate Assessment items. The use of classroom teachers from across the state for item writing is evidence of instructional validity pertaining to how well the test items reflect classroom instruction. Every year a diverse group of North Carolina educators is recruited to write items to replenish alternate item pools.

Standard 3.2 (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). Each new item undergoes a NCDPI iterative 16-step NCEXTEND1 Alternate Assessment item development and review process. Full details of the 16-step processes are documented in [Appendix 2–D \(p. 19–22\)](#).

The first two steps of the item development and review are mostly content focused. Upon receipt of newly written items, content specialists at NCSU-TOPS 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/EL/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 socioeconomic reasons. The specialists review the item’s assigned ECS, secondary ECS (if applicable), and key/appropriate foils. The specialist also reviews and make the following decision:

- If the content of the item is not accurate or does not match an objective or standard, 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, or graphs, 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.

At Step 4, two North Carolina-trained item reviewers review 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,
- Cognitive category,
- Being clearly written,
- Motivated and plausible distracters,
- Design conforming to North Carolina item writing guidelines,
- Appropriate language for the academic content area and age of students, and
- Bias or sensitivity concerns.

Steps 5 and 6 are reconcile of teacher content review and production edits. Step 7 is designed to address any potential accessibility issues and to ensure items are fair to all students. The EC, EL, and VI specialists review the item for accessibility concerns for the EC, EL, and VI students, such as accessibility of graphics for students with or without vision, and also consider accessibility in Braille. This review addresses concerns owing 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 socioeconomic reasons. Review also considered reading level of the item along with stem and foil quality—for example stem is a clear and complete question; foils are straightforward; no repetitive words; the grammar of the stem agrees with the foils; and idioms do not provide an accessibility issue.

All other items that either have no issues or had minor suggested reviews that were reconciled in Step 8 are forwarded to a second production edits for graphic (Step 9) and grammar review (Step 10). At Step 11, a security check is performed on all new items by production staff to make sure no duplicate copy of the item exists in the test development databases. If there is a duplicate copy of the item or a requested revision was not made, then the item is flagged and sent back to Step 8.

In Steps 12–14, items undergo final content and production reviews by alternate assessment lead (Step 12), NCSU-TOPS content lead review (Step 13), and final production review (Step 14). At Step 15, the TMS evaluates the item for alignment to grade level content standard; verifies that there is one and only one correct answer; cognitive category; bias,

sensitivity, or accessibility issues; and overall item quality. The TMS has these options after reviewing the item at Step 15:

- Approves the item as is, the item proceeds to Step 16 (item approved).
- Edits are needed, the item proceeds to Step 13 for review by a content specialist.
- Delete the item.

The item development and review process are continuous cycles to ensure sufficiency of the item pool. The finalized approved items are then field-tested and must undergo a post-field-test round of statistical reviews before they are placed on an operational form.

2.4 System Tryout Study

Since the *Edition 4* NCEXTEND1 tests are designed for primarily computer-based administration, a system tryout study was conducted in November–December 2017. NCDPI staff visited five North Carolina schools: four from Wake County and one from Johnston County. The teachers were recruited by their district. Each teacher selected a student who was eligible to take the *NCEXTEND1* NC Math 1 Alternate Assessment. The purpose of the system tryout was to assess user friendliness of the online system. This allowed NCDPI to understand firsthand the functionality of the online system and gain valuable feedback to help make improvements.

After administering the tryout to the student, NCDPI staff conducted a brief interview to gather teacher feedbacks. The feedbacks were implemented to enhance the online testing system. For details of the feedbacks please refer to [Appendix 2–E](#).

2.5 Mode of Test Administration

There are three (3) administration options available for the NCEXTEND1 Alternate Assessments. The IEP team determines, based on the individual needs of the student, which of the following options is most appropriate for the student:

- Option 1: teacher-facilitated online with the student recording responses on the computer/tablet (Optional: The paper test cards that coincide with the online test questions can also be placed in front of the student),
- Option 2: teacher-facilitated online with the teacher recording the student’s responses on the computer/tablet (Optional: The paper test cards that coincide with the online test questions can also be placed in front of the student.), or

- Option 3: teacher-facilitated online with paper test cards and the teacher recording the student’s responses on the computer/tablet.

2.6 *Edition 4* NCEXTEND1 Mathematics Alternate Assessment Field-Test Design for Grades 3–8 and NC Math 1

The main purpose of field testing prior to the development of new operational forms is to gather reliable item-level data to evaluate all aspects of item statistical characteristics, accessibility, and fairness, and to provide baseline statistical targets to assemble pre-equated parallel forms. Given the nature of the NCEXTEND1 student population, the use of standalone field-test administration have offered a flexible opportunity to gather essential item level data.

A standalone field-test design was conducted with seven forms and ten items in each form matrix sampled for grades 3–8 with a total of seventy items at each grade. Each item carried two-point. Hence each student can obtain a maximum of 20-points in each form. The NC Math 1 Alternate Assessment standalone field-test design included two forms with twenty-five items each with a total of fifty items. Since each item carried two-point, a student can score a maximum of 50 score points in each form. The forms were spiraled within each class, ensuring randomness of the student responses with cognitive complexities. A summary of the field-test plan is shown in *Table 2.4*. The rationale for standalone field-test design for the NCEXTEND1 assessment is the fact that number of students are smaller for embedding for stable item parameter estimates. This was done to protect the integrity and validity of students’ scores while comparing the student achievement with the general student population. Moreover, in subsequent administration the NCEXTEND1 tests are going to be primarily computer-based.

Table 2.4 NCEXTEND1 Mathematics Grades 3–8 and NC Math 1 Standalone Field-Test Plan, 2017–18

Grades	Field-Test Forms	Items in Each Form	Total Field- Test Items
3–8	7	10	70
10	2	25	50

Item tryout: During the field-test administration, a survey was conducted with assessors to provide feedback on NC Math 1 NCEXTEND1 test items’ appropriateness to the student population. The review was conducted in group setting with participation of all assessors in the school with one assessor entering and submitting the responses from the group. The school test coordinator designated the time and secure location of the review. Five to seven groups of

assessors were selected for reviewing ten items each. The assessors were asked the following questions:

- Question 1. Is there a correct answer?
- Question 2. Is the item question clear, concise, and written using simple language?
- Question 3. Is the vocabulary used in the item appropriate for this student population/grade/course?
- Question 4. Are the test card graphics appropriate?
- Question 5. Does the item content contain any fairness, sensitivity, or accessibility issues?
- Question 6. Is the overall item quality acceptable?

Most of the responses were positive that the test items were appropriate in terms of difficulty, content alignment, and sensitivity to their student population. Some of the typical comments included higher difficulty of the items for the population especially algebraic operation and exponential numbers, calculator active items that may be problematic for some group of students and confusing for others, and the items are academically towards higher end for the student population. These feedbacks were addressed when developing operational forms. Comments made by the assessors on the items across forms are presented in [Appendix 2–F](#).

CHAPTER 3 ITEM ANALYSIS

This chapter summarizes procedures and criteria NCDPI uses to analyze and evaluate the statistical and psychometric characteristics of newly developed NCEXTEND1 test items. Item analysis serves as the final quantitative process for item review and establishes grade level operational item pool for form development. Standard 4.10 (AERA, APA, & NCME, 2014) states, “*When a test developer evaluates the psychometric properties of items, the model used for that purpose should be documented. ... The process by which items are screened and the data used for screening, such as item difficulty, item discrimination, or differential item functioning (DIF) for major examinee groups, should also be documented*” (p. 89).

The NCEXTEND1 item analysis relied on Classical Test Theory (CTT) based statistics: percent correct (p-value), item-to-total correlations (biserial correlation), and distractor analysis to screen item quality following field tests. Item Response Theory (IRT) based statistics were not used primarily because the sample size per item was small, mostly less than 200, for evaluating IRT based item characteristics. The CTT procedures and flagging criteria used for item screening and analysis are explained and described in subsequent sections.

3.1 Statistical Item Flagging Criteria

The mathematics NCEXTEND1 Alternate Assessments are teacher-facilitated online assessments that are administered individually to each student where students have two trials with scaffolding to complete each item. The student who correctly answered the item in First Trial scores two-point. If the student misses the answer in the First Trial, the incorrect answer is removed from the choices for the Second Trial. The student who correctly answered the item in Second Trial scores one-point. The student who misses the answer in both trials scores zero.

The NCEXTEND1 tests are divided into two sets with Set 2 being relatively difficult than Set 1. The test terminates at the end of Set 1 for students who do not obtain enough scores. Students who obtained enough scores in Set 1 continues to Set 2. The assessment for the students ends after the last question in Set 2. All field-test items are classified into one of three NCDPI item flagging categories (Keep, Reserve, and Weak) based on trial one classical statistics with a goal to rank items in the final item pool for overall statistical quality. These specifications are routinely updated to continuously ensure that the highest quality items are selected for the NCEXTEND1 mathematics Alternate Assessments.

- **Keep:** These are items with good statistical properties from CTT procedures used for item analysis. Items flagged as “Keep” are first choice from the item pool during form assembly. Their main statistical properties are within the established NCDPI ranges considered as optimal items.

- Reserve: These are items with either p-value or point-biserial is barely outside the range to be considered as optional items. These items are only included in the final form assembly pool if they are needed to meet content or statistical specifications of the operational form. When any item flagged as “Reserve” from field tests is placed on a new form it must undergo additional content review to ensure the content is accurate.
- Weak: These are items with either p-value or point-biserial or both being significantly outside the range to be considered as optional items based on field-test analysis. When complete field-test data are available, these items are generally not included in the item pool used for form assembly. The only exception to this rule is when exceptional circumstances cause field-test data to be incomplete or unreliable. In such situations, thorough vetting is required from the content experts and psychometricians.

3.2 CTT Based Item Analysis

In accordance with NCDPI policy, whenever possible, all items must first be field tested prior to select them on operational form. After items are field tested, the first step for the NCEXTEND1 Alternate Assessment involved conducting a series of CTT analysis to determine if these items meet the minimum psychometric requirements to be considered for further evaluation. NCDPI uses a custom-developed SAS® Macro item analysis routine with a combination of procedures to process student response data from field tests and compute CTT statistics: item p-value, biserial correlation, and distractor analysis.

- Item p-value summarizes the proportion of examinees from a given sample answering the item correctly and is used as an indicator of item difficulty. Valid p-value for dichotomously scored items ranges between 0 and 1, where values close to 0 indicate extremely difficult items (few students selected the correct response) and values close to 1 indicate easier items (almost all students answered correctly).
- The polyserial correlation (Polycorr) coefficient is a special case of the Pearson correlation coefficient and describes the relationship between two continuous variables with a bivariate normal distribution. The biserial coefficient provides evidence of the strength of the relationship between the item and the unidimensional construct being measured. Theoretical range for biserial coefficient is -1 to 1 . Negative biserial correlation generally indicates the item might be measuring a separate unintended construct.
- Distractor analysis involves reviewing whether some distractors are frequently chosen over another by showing higher biserial correlation with the distractor.

Table 3.1 shows the CTT based item flagging criteria based on first trial.

Table 3.1 CTT Item Flagging Criteria

CTT Statistics	Flagging Criteria
$0.150 \leq p\text{-value} \leq 0.850$ and $\text{Polycorr} \geq 0.150$	Keep
$0.100 \leq p\text{-value} \leq 0.149$ or $0.851 \leq p\text{-value} \leq 0.900$ and $0.150 \leq \text{Polycorr} \leq 0.249$	Reserve
$p\text{-value} \leq 0.099$ or $p\text{-value} \geq 0.901$ and $\text{Polycorr} \leq 0.149$	Weak

Table 3.2 shows descriptive statistics of P-value and Point Biserial from the standalone field-test in 2017–18 for the NCEXTEND1 mathematics items. The statistics for the First Trial are based on responses when students answered the item for the first time. The possible scores for the First Trial are either 2 or 0. The statistics for the Second Trial are based on student responses from both trials. The possible scores in the Second Trial are 0, 1, and 2.

The descriptive statistics from *Table 3.2* indicated 90% or more items for NCEXTEND1 mathematics grades 3–8 and about 70% items for NC Math 1 are classified as meeting the NCDPI’s optimal standards of “Keep”. Moreover, the p–value and biserial ranges for both trials show the item pool had enough range of item difficulty and biserial correlation for high quality operational form assembly.

Table 3.2 CTT Descriptive Summary of Field-Test Item Pool, Spring 2018

Grade	Flag	Trial	Total Items	P-value				Point Biserial			
				Mean	SD	Min	Max	Mean	SD	Min	Max
3	Keep	First	63	0.46	0.11	0.22	0.81	0.43	0.10	0.20	0.58
		Second	63	0.65	0.17	0.22	0.88	0.44	0.09	0.18	0.60
	Reserve	First	2	0.35	0.05	0.32	0.39	0.16	0.05	0.12	0.19
		Second	2	0.35	0.05	0.32	0.39	0.23	0.06	0.18	0.27
	Weak	First	5	0.26	0.06	0.18	0.33	0.06	0.04	0.01	0.10
		Second	5	0.26	0.06	0.18	0.33	0.15	0.07	0.08	0.25
4	Keep	First	64	0.43	0.10	0.22	0.71	0.40	0.09	0.20	0.64
		Second	64	0.62	0.18	0.22	0.90	0.41	0.09	0.16	0.57
	Reserve	First	5	0.30	0.09	0.18	0.42	0.15	0.02	0.13	0.18
		Second	5	0.38	0.18	0.18	0.66	0.24	0.07	0.17	0.33
	Weak	First	1	0.21	.	0.21	0.21	0.06	.	0.06	0.06
		Second	1	0.21	.	0.21	0.21	0.21	.	0.21	0.21
5	Keep	First	68	0.41	0.09	0.24	0.67	0.40	0.09	0.20	0.60
		Second	68	0.59	0.18	0.24	0.86	0.42	0.08	0.22	0.59
	Reserve	First	2	0.25	0.01	0.24	0.26	0.14	0.02	0.12	0.15
		Second	2	0.43	0.24	0.26	0.59	0.27	0.01	0.26	0.28
6	Keep	First	67	0.40	0.10	0.20	0.69	0.38	0.08	0.22	0.59
		Second	67	0.61	0.19	0.20	0.88	0.39	0.08	0.13	0.59
	Reserve	First	3	0.29	0.08	0.22	0.39	0.17	0.01	0.16	0.18
		Second	3	0.29	0.08	0.22	0.39	0.23	0.05	0.18	0.27
7	Keep	First	70	0.43	0.11	0.22	0.69	0.39	0.09	0.22	0.61
		Second	70	0.61	0.19	0.24	0.88	0.40	0.07	0.24	0.58
8	Keep	First	66	0.38	0.09	0.21	0.66	0.37	0.09	0.19	0.54
		Second	66	0.57	0.19	0.26	0.88	0.37	0.08	0.12	0.58
	Reserve	First	2	0.39	0.08	0.33	0.45	0.15	0.01	0.14	0.16
		Second	2	0.53	0.11	0.45	0.61	0.26	0.01	0.25	0.27
	Weak	First	2	0.23	0.02	0.21	0.25	0.07	0.02	0.05	0.08
		Second	2	0.61	0.05	0.58	0.64	0.28	0.06	0.24	0.32
10	Keep	First	35	0.44	0.11	0.22	0.67	0.34	0.09	0.19	0.51
		Second	35	0.63	0.18	0.22	0.94	0.36	0.08	0.09	0.47
	Reserve	First	6	0.35	0.07	0.27	0.44	0.15	0.02	0.12	0.17
		Second	6	0.52	0.15	0.31	0.72	0.20	0.07	0.08	0.30
	Weak	First	9	0.29	0.08	0.21	0.46	0.08	0.02	0.02	0.10
		Second	9	0.40	0.15	0.21	0.62	0.15	0.09	0.02	0.31

3.3.1 Field-Test Sample Characteristics

Table 3.3 shows the demographic characteristics of the alternate population who attended the standalone field test including economically disadvantaged students (EDS) and ELs. It shows that the alternate population consisted of higher proportion of male and EDS. Moreover, the proportion of white students was the highest followed by black and Hispanic.

Table 3.3 Demographic Distribution (%) of 2017–18 Mathematics Field-Test Population

Grade	Gender		Ethnicity							EDS ELs		Total
	Female	Male	Asian	Black	Hispanic	American Indian	Multi-Racial	Native	White			
3	32.5	67.5	3.8	35.0	17.5	1.6	4.7	0.1	37.3	60.3	13.6	1,032
4	30.5	69.5	3.0	35.7	17.5	0.8	4.7	0.3	37.9	57.8	11.4	1,066
5	33.4	66.6	3.4	35.2	14.2	1.2	5.3	0.2	40.4	55.1	9.3	1,104
6	32.6	67.4	2.4	35.3	14.6	1.1	3.5	0.1	43.0	60.9	8.7	1,099
7	33.2	66.8	3.1	34.5	14.9	1.5	3.7	0.0	42.3	58.2	9.1	968
8	35.5	64.5	2.4	32.8	15.2	1.2	4.4	0.0	44.0	55.7	9.8	1,051
10	30.3	69.7	2.2	31.7	11.1	0.3	5.0	0.2	49.6	59.0	7.0	597

Table 3.4 shows the percentage of students by disability subgroups of the alternate student population who attended NCEXTEND1 standalone field test. It shows majority of the cognitively disable students were Autistic followed by Moderate Intellectual Disability and Mild Intellectual Disability as well as Multiple Disability.

Table 3.4 Distribution (%) of NCEXTEND1 2017–18 Mathematics Field-Test Student Population by Disability Categories

Disability Categories	Grades						
	3	4	5	6	7	8	10
Autism (AU)	43.5	40.0	38.2	35.0	33.6	36.6	32.3
Deaf-Blindness (DB)	0.2		0.2				
Serious Emotional Disability (ED)	0.5	0.1	0.4	0.5	0.1	0.4	0.3
Hearing Impairment (HI)	0.1	0.2	0.3	0.1		0.1	0.2
Intellectual Disability - Mild (IDMI)	13.7	15.3	13.9	15.6	15.7	13.2	13.9
Intellectual Disability - Moderate (IDMO)	20.8	23.5	21.5	25.8	25.4	28.3	31.0
Intellectual Disability - Severe (IDSE)	4.5	5.1	4.3	4.2	5.7	4.5	4.7
Specific Learning Disability (LD)	0.4	1.1	0.4	0.3	0.9	0.4	0.2
Multiple Disabilities (MU)	10.5	8.6	13.9	12.5	12.3	11.8	12.9
Other Health Impairment (OH)	4.2	4.6	5.3	4.3	5.3	3.4	3.7
Orthopedic Impairment (OI)	0.4	0.3	0.0	0.5	0.3	0.4	0.3
Speech or Language Impairment (SI)	0.1	0.1	0.1	0.0	0.0	0.1	0.0
Traumatic Brain Injury (TB)	1.3	1.1	1.6	1.2	0.7	0.8	0.5
Visual Impairment (VI)		0.0	0.1	0.2	0.0	0.1	0.0
Total	1,032	1,066	1,104	1,099	968	1,051	597

3.3.2 Raw Scores and Timing Data

Table 3.5 shows descriptive statistics of raw scores and time (minutes) taken by the students in the standalone field test. The grades 3–8 results indicate students on average took approximately one minute or less (9.5 to 10.4 minutes for the 10 items test) to complete each item across grades. Moreover, 95% of the student took slightly higher than one minute (14.3 to 17.2 minutes) to complete the 10-item test. Similar results were observed for grade 10 with average raw score of 22.7 and average time of 19.3 minutes taken by the students to complete the 25 items test. Ninety-five percent of the students completed the 25 items test form in 36.3 minutes.

Table 3.5 Grades 3–8 and NC Math 1 FT Descriptive Statistics of Raw Score and Timing Data

Grade	Sample	No. of Items	Raw Score					Time (Minutes)					
			Mean	SD	Median	Min	Max	Mean	SD	Median	Min	95th	Max
3	1046	10	10.4	3.8	11.0	0	20	8.0	9.2	7.0	1.8	14.3	281.4
4	1073	10	10.0	3.6	10.0	0	20	8.0	4.2	7.2	1.5	14.5	65.9
5	1111	10	9.8	3.7	10.0	0	20	8.4	3.6	7.7	0.9	15.7	27.8
6	1103	10	9.8	3.4	10.0	0	20	9.4	4.8	8.5	1.6	17.2	48.3
7	968	10	10.3	3.6	10.0	0	20	8.4	3.9	7.5	0.7	15.8	31.6
8	1045	10	9.5	3.4	9.0	0	20	8.9	4.5	8.1	1.2	16.0	48.7
10	599	25	22.7	7.8	24.0	0	41	19.3	9.1	17.3	4.7	36.3	79.4

3.3 Fairness Review

When constructing test forms, it is important to know the extent to which items perform differentially for various groups of students. The first step was flagging items for DIF. Since sample size for subgroups was small, no DIF analysis was conducted. The second step was convening a fairness review panel to examine all items. As a developer of the NCEXTEND1 assessments, it is the responsibility of NCDPI to examine all assessment items for possible sources of bias. The Standard 3.3 (AERA, APA, & NCME, 2014) states “*Those responsible for test development should include relevant subgroups in validity, reliability/precision, and other preliminary studies used when constructing the test*” (p. 64). Fairness is an ongoing concern when administering and constructing a summative statewide assessment. In order to meet this standard, NCDPI convened a Fairness Review panel. The same panel convened for content review were also used for the fairness review. These members were carefully selected based on their knowledge of the curriculum area and their diversity with respect to the student population. Prior to reviewing items, panelists had to complete an online fairness review training process through the NC Review System. The process is documented in [Appendix 3–A](#). Panelists were asked to evaluate the item based on the following questions:

- Does the item contain language that is not commonly used statewide or has different connotations in different parts of the state or in different cultural or gender groups?
- Does the item contain any local references that are not a part of the statewide curriculum?
- Does the item portray anyone in a stereotypical manner? (This could include activities, occupations, or emotions.)
- Does the item contain any demeaning or offensive materials?
- Does the item have offensive, stereotyping, derogatory, or proselytizing religious references?
- Does the item assume that all students come from the same socioeconomic background? (e.g., a suburban home with two-car garage)

- Does the artwork adequately reflect the diversity of the student population?
- Is there other bias or sensitivity concerns?

The online review platform requires that if there is any indication that the reviewer suspects an item is associated with a bias, sensitivity, or accessibility issue then he/she explicitly documents his/her concern. Following the review of all items by the panel, a final determination must be made whether to retain or delete any of these items from the operational item pool. Additional review was done by content test specialists at the NCDPI and NCSU-TOPS. These experts include, at a minimum, the Test Measurement Specialist, Psychometrician, and Lead Content Specialist at NCSU-TOPS. The less than optimal items were allowed to include on operational forms only if no other viable alternative is available in the item bank and all experts agree the items measured content that was expected to be mastered by all students and no obvious indication of specific construct-irrelevant variance is detected.

CHAPTER 4 OPERATIONAL FORM ASSEMBLY, ANALYSIS, AND REVIEW

AERA, APA, & NCME (2014) states, “*The test developer is responsible for documenting that the items selected for the test meet the requirements of the test specifications. In particular, the set of items selected for a new test form or an item pool for an adaptive test must meet both content and psychometric specifications*” (p. 82). To adhere to the standard, Chapter 4 documents the iterative IRT-based automated form assembly processes used to create parallel forms. This chapter also summarizes all the quality and content review steps the NCDPI used to finalize new operational base forms from the field-test pool. In all, the NCDPI has instituted a 21-step iterative form building and review process documented in [Appendix 2–D](#) (p.12–18).

4.1 Form Assembly and Statistical Targets of New Forms

The first step in form assembly requires the initial selection of items to match the test blueprint discussed in Chapter 2 and a statistical target for new forms. NCDPI uses a two-phase form assembly process to select and review forms. In Phase 1, an automated form assembly custom SAS® macro uses sampling procedures to optimally select items from the pool to match test blueprint and statistical specifications based on CTT statistics to recommend the most appropriate form.

Edition 4 mathematics NCEXTEND1 Alternate Assessments were developed to closely align with the test blueprint and average difficulty of the item pool. Note that the total number of items and test design of the *Edition 4* were changed from *Edition 3* Alternate Assessments. Attention was also paid to maximize measurement precision around the achievement level cuts at Not Proficient/Level 3 and Level 3/Level 4. These points are critical for reporting decisions of the NCEXTEND1 scales.

The statistical targets are determined independently for each grade based on the content complexity of grade level content standards and item statistics available in the item pool. The final statistical targets for base forms across grade are not intended to imply a vertical scale. *Table 4.1* shows the CTT based descriptive statistics of the operational forms based on field-test statistics. Note that the mathematics NCEXTEND1 tests are teacher administered by scaffolding of item options. The statistics for the First Trial are based on responses when students answered the item for the first time. The possible scores for the First Trial are either 2 or 0. The statistics for the Second Trial are based on student responses from both trials. The possible scores in the Second Trial are 0, 1, and 2.

Table 4.1 Field-Test CTT Summary Statistics, Spring 2018

Grade	N	First Trial		Second Trial	
		P-Value	Point Biserial	P-Value	Point Biserial
3	~166	0.45	0.43	0.66	0.45
4	~166	0.41	0.41	0.59	0.42
5	~166	0.42	0.38	0.61	0.40
6	~166	0.38	0.39	0.60	0.41
7	~166	0.39	0.36	0.55	0.39
8	~166	0.44	0.40	0.60	0.37
10	~300	0.45	0.31	0.61	0.35

4.2 Form Review

The NCEXTEND1 Alternate Assessments operational form review is a 24-step process (*Appendix 2–D, p. 25–28*). After the initial assembly and statistical review (Step 1) of the form development process is complete, the form then undergoes a series of iterative review process. At each critical review step, if there is a recommendation to replace an item the form is sent back to Step 1 of the test development for final consideration. If there is a replacement item from the bank that maintains the blueprint and statistical properties of the form, then a quick swap is made, and the form sent back through the review process.

4.2.1 Content Reviews

In the mathematics NCEXTEND1 operational form review processes, the content review steps are Steps 3–9 and Steps 14–20. These steps are conducted at various stages by a NCSU-TOPS content specialist, an NCDPI TMS, and an external outside content reviewer. The ultimate objective of content review is to make sure all items selected on tests are appropriate and aligned to grade level content. They also check to make sure items on forms do not cue and are not repetitive (for example overemphasis on a subtopic and if all area problems in one form were isosceles triangles). Criteria for evaluating each test form included the following:

- The content of the test forms reflects the goals and objectives of the North Carolina ECS for the subject (content validity).
- The content of test forms reflects the goals and objectives as taught in North Carolina schools (instructional validity).
- Items are clearly and concisely written, and the vocabulary is appropriate to the target age level (universal design).

- Content standards of the test forms are balanced, and items do not cue other items on a form.
- All selected response items have one and only one best correct response choice. The distractors should appear plausible for someone who has not achieved mastery of the representative objective (one best answer).

The outside content reviewers are instructed to complete a mock administration of a test form and to provide written comments and feedback next to each item. Each reviewer independently documented his or her opinion as to how well the tests met the five criteria listed above. These comments were further reviewed by NCSU-TOPS and NCDPI content specialists with the goal to address concerns ranging from a simple grammatical fix to replacing the item from the form.

At Step 20, a content manager reviews comments/suggestions 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 22 as approved.
- 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 21 (Production Edits) for revisions to artwork, graphs, or ELA selections.
- Reject the form.

4.2.2 Production Reviews

Production and grammar reviews of text, artwork or graphs, and copyright are continuously monitored and checked in several steps (Steps 2, 4, 7, 10, 12, 15, 18 and 21) of the 24-step processes. Most of the production steps are used for revision of items such as minor grammatical edits, formatting and revision of artwork or figures on items. All proposed revisions to base form items must be approved by the psychometrician who will determine if proposed edits are significant to the point that it might affect the interpretation of field-test statistics. If it is ruled the proposed revision will invalidate the item field-test statistics, then a recommendation is made to replace the item.

At Step 21, revisions to items such as artwork and graphs selections are made by production staff. Once the revisions are made, the form is sent back to Step 20 final Manager review. Any suggestions that are rejected should be noted in the form comments. Any suggested edits to operational items that Content Staff feel warrant consideration are directed to the TMS and Psychometrician for consideration.

After final review of the PDF online, the computer-based forms are exported from the TDS application into the NCTest Admin platform. In this stage, a series of quality checks are performed by staff to ensure all the specified interactions between items and the NCTest Admin are fully functional across different end users' approved browsers.

Each form is assigned to a demo student so the forms may be reviewed by the NCDPI's test measurement specialists. This is to ensure the images and text display properly on the screen.

All forms are sent to Step 24 as approved, and all items are operationally locked to prevent any further revisions. This is to ensure that the published versions of the form, items and selections are preserved electronically.

4.2.3 Bias and Sensitivity Reviews

There are several bias review processes built into the development of the mathematics NCEXTEND1 items and forms that are intended to prevent content with bias and sensitivity issues. These processes begin on an item development level with all test development specialists, item writers, and reviewers trained on the principles of universal design and best practices in assessment. This training includes guidelines on how to create items that are fair and reflect the diversity of North Carolina's student population with special attention given to socio-economic status, culture, and language considerations.

Teacher 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

Exceptional Children (EC), English as a Second Language (ESL), and Visually Impaired (VI) Reviewers review 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 or without vision and also considers brailing accessibility. This review addresses concerns due to bias or sensitivity 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 including:

- 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.

Detail bias and sensitivity review processes are documented in [Appendix 3–A](#). In 2017–18 administration, the Fairness Review panel for both general and NCEXTEND1 mathematics was made up of 12 participants representing teachers and educators. These members were selectively recruited based on their expert knowledge of mathematics content. Their demographic information is summarized in [Table 4.2](#). Prior to reviewing items, panelists had to complete a training on the fairness review process online through the NC Review System.

Table 4.2 Demographic Information for Fairness Review Panels, Spring 2018

Category	Subcategory	N	%
Gender	Female	4	33%
	Male	8	67%
Ethnicity	Black	1	8%
	White	11	92%
Highest Degrees Earned	J.D./Ed.D./Ph.D.	8	67%
	MA/MS/M.Ed.	4	33%
Year of Experience	>20	3	25%
	10–20	7	58%
	1–10	2	17%

4.3 Summary of Final Operational Forms and Field-Test Design

This section details test structures and statistical properties of new *Edition 4* NCEXTEND1 mathematics alternate test forms that were built in 2018 using items from standalone field-test. All forms were built based on test specification criteria outlined in Chapter 2.

4.3.1. *Edition 4* NCEXTEND1 Mathematics Test Structures and DOKs

Table 4.3 displays test structure of the final assembled operational base forms in terms of number of items, item types, answer options, score per item, and DOKs. Each item contains three distractors and carry a maximum of 2-point if the student answered correctly in first trial and 1-point if the student answered correctly in second trial. The table also shows DOK levels in terms of percentage.

Table 4.3 Test Structures, Grades 3–8, and NC Math 1

Grade	Number of Items			Item Type	Answer Options	Score Per Item	DOK Level (%)	
	Operational	Field-Test	Total				1	2
3	27	5	32	SR	3	2	60-70	30-40
4	27	5	32	SR	3	2	60-70	30-40
5	27	5	32	SR	3	2	60-70	30-40
6	27	5	32	SR	3	2	50-60	40-50
7	27	5	32	SR	3	2	50-60	40-50
8	27	5	32	SR	3	2	50-60	40-50
10	28	5	33	SR	3	2	60-70	30-40

SR=Selected-Response

4.3.2. Field-Test Plan

NCEXTEND1 NC Math 1 operational test contained 28 operational items and six (6) slots for field-test items. Similarly, grades 3–8 contained 27 operational items with five (5) slots for field-test items. The items are presented to students in a Two-Set fixed format design. Two sets are designed to address the issue that students who are not responsive to the test do not have to sit there for whole time. Set 1 contained a total of 22 selected-response (SR) items in NC Math 1 and 20 SR items in grades 3–8 with 3 response foils. Students will have up to two trials to select the correct response. A correct response in First Trial will be awarded two-point and one-point if the correct response is selected in Second Trial.

Set 2 contained a total of 12 SR items in NC Math 1 and 10 SR items in grades 3–8 with students given up to two trials to select the correct response. The same scoring rule apply; two points for correct response in First Trial and one point for correct response in Second Trial.

All students who earn at least five points of the possible 34 total score points available for grades 3–8 and 36 for NC Math 1 in Set 1 advanced to complete Set 2 items. For students with a score of four or lower from Set 1, their test session ended at the conclusion of Set 1. Scaffolding applied to all items in both Set 1 and Set 2 in the Second Trial, if the students did not respond or

select the correct response in the First Trial. The incorrect foil the student selected in the First Trial were removed and the item presented again in Second Trial with the two remaining foils. All items are scored as follow; 2-point for correct response in First Trial and 1-point for correct response in Second Trial. The design of the field-test plan is shown in *Table 4.4*.

Table 4.4 Grades 3–8 and NC Math 1 Field-Test Design

Modules	Operational		Field Test		Total
	Items	Points	Items	Points	Items
NC Math 1					
1	18	36	4	8	22
2	10	20	2	4	12
Total	28	56	6	12	34
Grades 3–8					
1	17	34	3	6	20
2	10	20	2	4	12
Total	27	54	5	10	32

A single base form with two flavors were administered during each administration cycle at every grade level. Each flavor consisted of the same operational but different field-test items. Flavors were randomly assigned to each grade classroom within school. This ensured each teacher administered a single form and flavor combination of items to all students in their classroom. The goal during form-flavor assignment is to have about 50% of classroom/students being administered each flavor. At the conclusion of every testing cycle, the plan is to evaluate the base form to possibly replace some of the operational items with recently field-test items while maintaining the psychometric scale.

CHAPTER 5 TEST ADMINISTRATION

Standard 6.0 (AERA, APA, & NCME, 2014) states, “*To support useful interpretations of score results, assessment instruments should have established procedures for test administration, scoring, reporting, and interpretation. Those responsible for administering, scoring, reporting, and interpreting should have sufficient training and supports to help them follow the established procedures...*” (p.114). In adherence to this standard, this chapter briefly describes NCDPI’s established policies and procedures used to train test coordinators and assessors in order to ensure mathematics grades 3-8 and NC Math 1 NCEXTEND1 standardized test administrations across the state. This chapter also provides information about NCEXTEND1 test administration guides, testing windows, mode of administrations, timing guidelines, testing accommodations and mechanism for reporting test irregularities and misadministration.

5.1 Test Administration Guides and the Test Coordinators’ Handbook

The NCDPI produces a comprehensive Alternate Assessment guide for the mathematics grades 3-8 and NC Math 1 NCEXTEND1 Alternate Assessment. The guides for assessors and coordinators are available to ensure standardized administration of all NCEXTEND1 Alternate Assessments given across the state. They are briefly described below with website links for more detailed descriptions.

NCEXTEND1 Alternate Assessment Guide: The mathematics grades 3-8 and NC Math 1 NCEXTEND1 Alternate Assessments measure the ECS adopted by the NCSBE in 2017. The assessments are administered online to all eligible students at the appropriate grade levels with paper/pencil administration as an option. The NCEXTEND1 Alternate Assessment Guide contains the administrative procedures that apply to both online and the paper/pencil administrations. Any information that is specific to either the online or paper/pencil format is clearly labeled for assessors. The guide is for test administrators to administer the tests and is internal to the NCDPI.

Testing Students with Disabilities: This document contains policy guidelines and procedures for testing students with disabilities in the North Carolina Annual Testing Program including students with significant cognitive disability. Students who are classified as having a disability or a significant cognitive disability are those who receive special education and related services under the *Individuals with Disabilities Education Improvement Act of 2004 (IDEA)* or identified with a disability under *Section 504 of the Rehabilitation Act of 1973*. School personnel must ensure the policy guidelines and procedures outlined in this publication are implemented appropriately. The policy guidelines and procedures for testing students with disabilities can be found in [Appendix 5–A](#).

Testing Security Protocols and Procedures for School Personnel: The NCDPI publishes this document in order to maintain the integrity of the North Carolina Annual Testing Program for the general as well as for the alternate student population. It is essential for school personnel to develop awareness of proper testing protocol and procedures. Knowledge of testing policies and procedures helps ensure the North Carolina Annual Testing Program for the students with significant cognitive disability to conduct in a manner that is fair, consistent and equitable for all students. The purpose of this publication is to provide principals, teachers and other school personnel with a reference for implementing secure, uniform test administrations for the North Carolina Annual Testing Program. This testing security guide may be kept in the schools. The document link can be accessed in [Appendix 5–B](#).

North Carolina Test Coordinators’ Policies and Procedures Handbook: The purpose of this Handbook is to provide Local Education Agency (LEA) and charter school test coordinators with a reference for implementing proper test administrations for the North Carolina Annual Testing Programs. This handbook provides information to ensure that the integrity of the testing program is maintained, results generated from the program are valid, and any subsequent reporting is accurate and appropriate. The document link can be accessed in [Appendix 5–C](#).

5.2 Alternate Assessment Eligibility Criteria

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

- The student must have a current Individualized Education Program (IEP).
- The student must have a significant cognitive disability:
 - The student’s disability significantly impacts cognitive functioning and adaptive behaviors, defined as those skills which are essential for someone to live and function independently.
 - The student requires extensive and repeated individualized instruction and support to make meaningful gains.
 - The student uses substantially adapted materials and individualized methods of accessing information in alternative ways.
- The student must be instructed using the North Carolina Extended Content Standards (i.e., reading and mathematics) and the North Carolina Extended Essential Standards (i.e., science).
- The student must be enrolled in grades 3–8, 10, or 11, according to PowerSchool. Only those students enrolled in 11th grade for the first time are required to take the NCEXTEND1 alternate assessment at grade 11.

Most 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 the assessed general grade- or course-level content standards of the North Carolina Standard Course of Study (i.e., reading, mathematics) and the Essential Standards (i.e., science),
- demonstrate delays only in academic achievement,
- demonstrate delays only in selected areas of academic achievement,
- demonstrate delays owing primarily to behavioral issues, or
- if in high school, are pursuing a North Carolina high school diploma (including students enrolled in the Occupational Course of Study pathway).

Evidence for the decision to participate in NCEXTEND1 is not based on:

- a disability category or label,
- poor attendance or extended absences,
- native language, social, cultural, or economic differences,
- expected poor performance on the general education assessment,
- academic or other services the student receives,
- educational environment or instructional setting,
- percent of time receiving special education services,
- English Learner status,
- low reading level or achievement level,
- anticipated disruptive behavior,
- impact of student scores on the accountability program,
- administrative decisions,
- anticipated emotional distress; or
- need for accommodations to participate in the assessment process.

IEP teams may use the North Carolina Alternate Assessment Decision Making Flow Chart ([Appendix 5-D](#)) to aid in decision making regarding the NCEXTEND1 alternate assessment for students.

5.3 Assessor Training

The assessors' training leading to the testing day utilizes the *North Carolina Test Coordinators' Policies and Procedures Handbook*, the NCEXTEND1 Alternate Assessment guide as well as all other NCDPI publications discussed in Section 5.1. These documents contain comprehensive information on test administration including test security, roles and responsibilities of test administrators, test administration preparation, monitoring, testing accommodations, online testing, testing irregularities and available resources.

On a day before each mathematics NCEXTEND1 test administration, all assessors must be trained in test security and testing procedures. The school system or school test coordinator will schedule and conduct the training session. Assessors shall:

- Prepare for and attend assessor training session(s) each testing cycle before administering any secure state assessment.
- Read and sign a Test Security Agreement (provided by the school test coordinator) at the conclusion of every test administration training session (All signed Test Security Agreements must be kept on file with the school test coordinator.)
- Read the assessment guide thoroughly.
- Review the sample assessor script(s) in the assessment guide and be prepared for the variations required by the testing conditions.
- For online mathematics testing, complete the Online Practice Activity; and
- For online mathematics testing, review other technical Issues.

To be made aware of appropriate use of accommodations, an assessor who will be conducting test administrations with accommodations must also be trained by the school system test coordinator or designee in the provision of the specified accommodations before the test administration.

5.4 Test Security and Administration Policies

Test security is an ongoing concern for the North Carolina Annual Testing Program. When test security is compromised, it can undermine the validity of test scores. For this reason, NCDPI has taken steps to ensure the security of the assessments by establishing protocols for school employees administering tests. The test security guidelines that should be discussed during the test administration training include.

- *Copying Secure Test Materials:* No person may copy, reproduce, or paraphrase the test materials in any manner for any reason without the prior written consent of the NCDPI.
- *Classroom Instruction and/or Study Guides:* Secure tests must **not** be copied, filed, or used at any time during classroom instruction or in resource materials such as study guides. Teachers are **not** permitted to discuss specific items from the tests with students or colleagues before, during, or after the test administration or to ask students which test questions were difficult.
- *Testing Environment:* All rooms designated for testing, including rooms to which students may be relocated, must be quiet, orderly, and comfortable with adequate seating, lighting, and heating/cooling.
- *Accounting for and Storing Test Materials:* The assessor must count and record the number of secure test materials and supplemental materials. For online testing, devices that are either open and display the students' start screens or test material or are paused must not be left unattended by the assessor at any time.

5.4.1 Protocols for Assessors

Only school system employees are permitted to administer secure state tests. Those employees must participate in the training for test administrators as described in Section 5.2. Test administrators may not modify, change, alter, or tamper with student responses. Test administrators must thoroughly read and be trained on the appropriate *Test Administration Guide* and the codified North Carolina *Testing Code of Ethics* prior to the test administration. Test administrators must follow the instructions 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.

5.4.2 Protocol for Handling Alternate Assessments

The test materials include a packet of paper test cards that coincide with each computer-based item. Every assessor receives a packet of test cards for each student that they will access. The test cards indicate the test form the student will be administered. Use of the test cards is optional for assessors during the actual test administration. The decision to use the cards is based on the individual needs of the student.

The assessor script (i.e., student directions/test questions/answer choices) displays on the device screen. During the actual assessment, the script must be read directly from the device screen. Sample scripts of the student directions are provided as a training resource for assessors in Appendices C through H of the *NCEXTEND1 Alternate Assessment Guide*. There are three (3) administration options available for the mathematics NCEXTEND1 Alternate Assessments. The Individualized Education Program (IEP) team must determine, based on the individual needs of the student, which of the following options is most appropriate for the student:

- Option 1: teacher-facilitated online with student recording responses on the device (Optional: The paper test cards that coincide with the online test questions can also be placed in front of the student.).
- Option 2: teacher-facilitated online with teacher recording responses on the device (Optional: The paper test cards that coincide with the online test questions can also be placed in front of the student.) or
- Option 3: teacher-facilitated online with paper test cards and teacher recording responses on the device.

If paper test cards are used and presented to a student during an administration, the cards *must* match what is displayed on the device screen.

5.5 Test Administration

Standard 6.1 (AERA, APA, & NCME, 2014) states, “*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). The standardized procedures reduce construct-irrelevant variance and enhance the reliability and validity of the resulting test scores. The following subsections briefly describe testing windows and mode of test administration.

5.5.1 Testing Windows

The testing window for the mathematics NCEXTEND1 Alternate Assessments at grades 3–8, and NC Math 1 is the last ten (10) days of the school year. For school systems that were required to adjust their school schedules because of adverse weather conditions or other emergencies, the testing schedule should be adjusted to fall within the last ten (10) days of the adjusted school year. Exceptions are permitted to allow testing of a student outside the designated testing window to accommodate a student’s IEP as well as in rare circumstances where family emergency, family relocation, and scheduled surgery during the test window.

5.5.2 Modes of Test Administration

All grades 3–8 and NC Math 1 mathematics NCEXTEND1 tests for 2018–19 and beyond are computer-based with three answer recording options as stated in Section 5.4.2. Before the administration of the NCEXTEND1 Alternate Assessments, teachers must complete one of the answer recording option practice activities pertinent to the student.

Online Practice Activities

The practice activities assist teachers in determining which administration option is most appropriate for their students. Additionally, students participating in the NCEXTEND1 Alternate Assessments should complete each content area practice activity at least one time at the school before test day. The practice activities can help students become familiar with the testing platform and practice responding to sample test items. For best results, students should complete the practice activity using the device they will use during the actual assessment. It is, however, not necessary for students to complete the practice activity if they will be administered the assessment using just the test cards with the assessor recording the responses on the device.

Test Administration Preparation

In order to prepare for the assessment, the assessor must follow the *NCEXTEND1 Alternate Assessment Guide*. The assessor must have a device meeting all technical requirements. Technical Requirements for NCTest Admin--one grade-appropriate test card kit per student assessed, a four-function calculator with memory key (for grades 6–8 and NC Math 1 only), and a supply of scratch paper and sharpened pencils with erasers or other student-specific writing utensils.

Testing Schedule for Online Assessments

Online assessments have the following scheduling options:

- Online assessments should be administered as early in the school day as the school schedule permits; however, afternoon administrations are also permissible.
- For best practices, the North Carolina Testing Program strongly recommends schools schedule each online assessment within the school day for the maximum time allowed without interfering with lunch or bus schedules. All students, including students with the Scheduled Extended Time accommodation, must be allowed to eat lunch during the school day and to travel home at their regularly scheduled time.
- Schools may elect to administer two assessments in one day. However, caution must be taken when scheduling such testing to ensure that students can eat lunch during the school day and travel home at their regularly scheduled time.

5.6 Technical Specifications for NCEXTEND1 Assessment

To ensure students receive valid and reliable assessment administrations, schools must meet specific technical requirements. The devices that will be used for the online mathematics test administrations should meet the specified technical requirements. Schools must review these technical requirements on days before an online assessment and must make any necessary adjustments before administering an online assessment. The NCTest Admin is used as a secure site for the NCEXTEND1 mathematics assessment administration. In the event of technical difficulties during the actual test administration, the assessor must contact the school test coordinator. Additionally, assessors have the following responsibilities:

- Ensure NC Education usernames and passwords are working correctly, and access is available to the test's student information pages. Usernames and passwords should be checked two to three days before the test administration date.
- Check that all students who are to participate in the online assessment are listed in the Enrollment Tab list for the specific assessment in NCTest Admin.

- Ensure the test administration sessions are scheduled in NCTest Admin. Test administration sessions must be scheduled no later than the day before the test administration date.

5.7 Testing Time Guidelines

The AERA, APA, & NCME (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). The grades 3-8 and NC Math 1 mathematics 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 mathematics NCEXTEND1 Alternate Assessments may be administered over several days or may be completed in one session. 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. *Multiple Testing Sessions* is most appropriately used when a student is purposefully scheduled to take the assessment in specifically timed increments (e.g., three items a day over five days, 15-minute testing sessions, etc.). Otherwise, built-in breaks identified in the assessment design are most appropriate. The test design for the mathematics NCEXTEND1 Alternate Assessments allows breaks to be taken at any time during testing if the need arises, regardless of documentation in the student’s IEP. The assessor must use professional judgment to determine when a break is needed and what is an appropriate length of time for a student’s test administration. All test materials shall remain secure during all breaks.

Summary timing data for the 2018–19 NCEXTEND1 operational mathematics assessments are shown in *Table 5.1*. Extreme data including multiple testing days and negative time data were excluded from computation. The average time taken by the students to complete the operational test is about 36 minutes or less—about one minute per item. Moreover, 95% of students completed the assessments closely within 1.5 hours (96 minutes or less) window. Notice that the standard deviation (SD) is large indicating high variation of examinees taking time to complete the tests. Some students took longer (3 hours or more), and it was not clear if it was the actual time students took or it was teachers who took more time to enter the responses.

Table 5.1 Recorded Test Duration (Minutes) for Grades 3–8 and NC Math 1 Operational Forms, 2018–19

Grade	N	Number of Items (OP+FT)	Statistics		Percentile				
			Average	SD	25th	Median	75th	95th	99th
3	1,231	27+5=32	32	36	18	21	30	93	222
4	1,266	27+5=32	31	33	18	21	30	78	219
5	1,278	27+5=32	32	32	18	24	33	81	198
6	1,292	27+5=32	36	38	21	27	38	96	231
7	1,327	27+5=32	33	30	18	27	36	75	183
8	1,249	27+5=32	33	36	18	24	33	84	222
10	1,013	28+6=34	33	31	18	27	36	78	183

5.8 Testing Accommodations

State and federal law requires that all students, including students with significant cognitive disabilities and students identified as ELs, participate in the statewide testing program. Students with significant cognitive disabilities may participate in the standard NCEXTEND1 Alternate Assessments on grade level as it is designed with or without testing accommodations. AERA, APA, & NCME (2014) states that the eligible students participating in the NCEXTEND1 assessments are provided with “*test accommodations, when appropriate and feasible, to remove construct-irrelevant barriers that otherwise would interfere with examinees’ ability to demonstrate their standing on the target constructs*” (p. 67). Shyyan et al. (2016) define testing accommodations 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*”.

Accommodations are provided to eligible significant cognitive disability students with appropriate administrative procedures to assure that individual student needs are met while maintaining sufficient uniformity of the test administration. Approved accommodations for the grades 3-8 and NC Math 1 mathematics NCEXTEND1 Alternate Assessments for students with current IEPs are shown in Table 5.2.

Table 5.2 Approved accommodations for the NCEXTEND1 Alternate Assessments

Accommodations	Reading Grades 3–8 & 10	Mathematics Grades 3–8 & 10	Science Grades 5, 8, & 10
Braille Materials (Full UEB)	Yes	Yes	Yes
Braille Writer/Braille Paper	Yes	Yes	Yes
Large Print Materials	Yes	Yes	Yes
Assistive Technology Devices	Yes	Yes	Yes
Interpreter/Transliterators/Signs/Cues Test	NCEXTEND1 tests are to be read aloud to all students as specified in the NCEXTEND1 Assessment Guide		
Magnification Devices	Yes	Yes	Yes
Word-to-Word Bilingual (English/Native Language) Dictionary/Electronic Translator (ELs only)	Yes	Yes	Yes
Test Read Aloud (in English)	NCEXTEND1 tests are to be read aloud to all students as specified in the NCEXTEND1 Assessment Guide		
Multiple Testing Sessions	Yes	Yes	Yes
Testing in a Separate Room	Yes	Yes	Yes
Adaptations to NCDPI-Provided Manipulatives	Yes	Yes	Yes
Special NCDPI approved accommodation(s)			

For detail information about the approved accommodations for the grades 3-8 and NC Math 1 mathematics with current IEPs visit to testing students with disability guidelines described in the *NCEXTEND1 Alternate Assessment Guide*.

5.9 Student Participation

For a student with significant cognitive disability to participate in a mathematics NCEXTEND1 Alternate Assessment, the student must meet the eligibility criteria established by the NCDPI shown in Section 5.2, and the decision to participate in the Alternate Assessment must be documented in the current IEP. The *Individuals with Disabilities Education Improvement Act of 2004 (IDEA)* and *Every Student Succeeds Act (ESSA)* require students with disabilities to participate in an Alternate Assessment, with or without accommodations, if they are receiving instruction through the North Carolina ECS. The eligibility criteria in Section 5.2 indicates only

students with the most significant cognitive disabilities may participate in the Alternate Assessment, and no student can take the Alternate Assessment unless the student received classroom instruction in the North Carolina ECS. In North Carolina, a student with the most significant cognitive disability is defined as a student whose disability significantly impacts adaptive behaviors, defined as those skills which are essential for someone to live and function independently, and who;

- requires extensive and repeated individualized instruction and support to make meaningful gains, and
- uses substantially adapted materials and individualized methods of accessing information in alternative ways.

Under ESSA, the number of students who may take the Alternate Assessment is limited to no more than 1% of the total number of all students in the state who are assessed in a given subject (i.e., reading, mathematics, and science). The 1% threshold is at the state level. ESSA further requires that a school district exceeding 1% participation in any subject must submit a justification to the department, and the department must provide appropriate oversight to that district. The state and districts must still meet the 95% participation rate for students enrolled in tested grades. The percentages of students who participated in the 2018–19 mathematics general and the NCEXTEND1 Alternate Assessment are presented in *Table 5.2*. The NCEXTEND1 students make up of approximately 1% (rounded) of the total student population assessed.

Table 5.3 Participation of Students (%) by Assessment – General and NCEXTEND1

Grade	General		NCEXTEND1		Total N
	N	%	N	%	
3	116,059	98.9	1,275	1.1	117,334
4	120,320	98.9	1,312	1.1	121,632
5	121,935	98.9	1,325	1.1	123,260
6	121,613	98.9	1,332	1.1	122,945
7	118,471	98.9	1,367	1.1	119,838
8	80,897	98.5	1,270	1.5	82,167
10	117,996	99.1	1,059	0.9	119,055

5.10 Medical Exception

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. The medical emergencies may include, but not limited to, circumstances involving students who are i) in the final stage of a terminal or degenerative illness, ii) comatose, or iii) receiving extensive short-term terminal treatment due to

a medical emergency. For requests that involve significant medical emergencies and/or conditions, a school may request from the Division of Accountability Services/North Carolina Annual Testing Program a testing exception for the student. There is a process in place for requesting the medical exception. The request must be submitted electronically through a secure portal and with the superintendent’s or school director’s permission. The request must include detailed justification explaining why the student’s medical emergency and/or condition 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 electronic documentation containing identifiable student information is not disseminated or otherwise made available to the public. For more information on the process for requesting medical exceptions based on significant medical emergencies and/or conditions, please access the annual [Request for Testing Exceptions Based on Significant Medical Emergencies and/or Conditions](https://files.nc.gov/dpi/documents/files/med-exception-memo_ds_th_081820.pdf). https://files.nc.gov/dpi/documents/files/med-exception-memo_ds_th_081820.pdf.

5.11 Test Irregularity and Misadministration

Standard 6.7 (AERA, APA, & NCME, 2014) states, “*Test users have the responsibility of protecting the security of test materials at all times*” (p. 117). Any action that compromises test security or score validity is prohibited. These may be classified as testing irregularities or misadministration. NCDPI has a process in place to report testing irregularities and misadministration. A sample test security action plan is shown in the *North Carolina Test Coordinator Policies and Procedures Handbook* (Appendix 5–C).

School systems must monitor test administration procedures. According to *NCSBE policy TEST-001*, if school officials discover any instance of improper administration and determine the validity of the test results has been affected, they must notify the local board of education, order the affected students to be retested, and declare a misadministration. Only the superintendent and the school system test coordinator have the authority to declare misadministration at the local level. When a misadministration is declared, the affected student(s) must have the secure form of the NCEXTEND1 Alternate Assessment readministered. The public school unit (PSU) will specify how misadministration is to be handled at the school and will schedule dates and times for readministering the tests in each school. Only scores resulting from a valid test administration should be included in students’ permanent records, used for placement decisions, or used for accountability purposes. All misadministration must be documented and reported using the appropriate procedures outlined in the Online Testing Irregularity Submission System (OTISS).

Details regarding what constitutes a testing violation or irregularity and information concerning how to report incidents should be discussed during assessor training. Assessors must report any alleged testing violation or testing irregularity to the school test coordinator on the day of the occurrence. Examples of testing irregularities include, but are not limited to, the following. For further details refer to *online testing irregularity submission system* document (*Appendix 5–E*)

- i) Eligibility Issues:
 - Eligible students were not tested.
 - Ineligible students were tested.
- ii) Test Administration Issues:
 - Teacher/proctor failed to follow directions
 - Approved accommodation/exemption not provided
 - Approved accommodation not provided appropriately
 - Accommodation “read aloud” used in reading
 - Accommodation/exemption used but not approved/documented
 - School staff provided materials improperly
 - Student not required to take a test
 - Failure to test eligible students
 - Defective test materials
 - School staff administered the wrong test
- iii) Test Security Issues:
 - Failure to remove inappropriate displays
 - Secure material divulged
 - Teacher/proctor inadequately supervised testing
 - Improper use of test materials
 - Missing test material
 - Test books not properly returned
 - Test materials not stored in secure locked area
 - Items from secure test used for instruction
 - Reproduction of secure test in any manner or form
 - Making copies of test available to others
 - Online test connectivity/technical problems
 - Failure to delete secure electronic files
 - Incorrect/wrong number of materials given to school
- iv) Student Related Issues:
 - Student(s):
 - cheated by copying, cheat sheet, asking for info.
 - ill/had injury
 - was anxious
 - had a problem with medication, glasses, etc.

- absent
- went to bathroom
- left school for appointment
- refused to take test
- marked test randomly
- fell asleep
- did not pay attention
- was disruptive
- misaligned answer sheet
- did not follow directions
- worked on wrong section of test
- used calculator inappropriately
- brought inappropriate resource
- has family/personal problem
- External noises/disruption
- refused to use approved accommodation
- Fire alarm/bomb threat present
- v) Procedural Issues:
 - Encouraging students to be absent
 - Modifying test directions for standard administration
 - Teacher altered responses
 - Teacher/proctor gave improper assistance or provided improper instruction
 - Test not administered on designated date/window
- vi) Other Issues:
 - Cell phone

Note that schools must report online test connectivity and technical problems that occur during the administration of online assessments when a student is not able to successfully complete the assessment. Reports do not need to be entered for students who successfully complete the assessment despite a technical issue. If the same technical problem is being reported for multiple students for the same test administration on the same day, only one OTISS report needs to be submitted. A list of all students affected should be attached to the OTISS report.

5.12 Data Forensics Analysis

Maintaining the validity of test scores is essential in any high-stakes assessment program and misconduct represents a serious threat to test score validity. When used appropriately, data forensic analyses can serve as an integral component of a wider test security protocol. The results of these data forensic analyses may be instrumental in identifying potential cases of misconduct

for further follow-up and investigation. The possible data forensics analyses on NCDPI's operational assessments included:

Longitudinal Performance Comparison. NCDPI psychometricians compare longitudinal performance in terms of mean scale scores and proportion of students in different achievement levels on the grades 3-8 and NC Math 1 mathematics NCEXTEND1 Alternate Assessments across test administrations. Any unusual performance gains may be indicative of possible irregularity issues and may suggest of further exploration.

Testing Outside of the Window Monitoring. Schools are monitored to ensure that all state testing is completed within the state-mandated testing window. NCDPI has established set dates and windows for all state required testing. If testing occurs outside of the mandated testing window, the school must submit an irregularity report in OTISS.

CHAPTER 6 SCORING AND SCALE DEVELOPMENT

This chapter describes procedures used by the NCDPI to collect, certify, and score the grades 3–8 and NC Math 1 mathematics NCEXTEND1 Alternate Assessment student responses to create final reportable scale scores. The NCDPI uses linear transformation of raw scores for reporting scale. The procedures and steps, described in subsequent sections, are used to ensure student response data are securely and reliably scored so uses and interpretation of mathematics NCEXTEND1 Alternate Assessment scale scores are valid and fair for all students across the state.

6.1 Scoring and Scale Scores

Due to relatively small Alternate Assessment population and likely small variation of abilities, use of IRT models was not appropriate for the assessments. Instead, raw scores were linearly transformed to reportable scale scores using targeted mean and standard deviation as scaling constants. A pre-established scoring model has been traditionally used in North Carolina beginning in early 1990s and remained an important feature in the grades 3–8 and NC Math 1 mathematics NCEXTEND1 Alternate Assessments. The use of the pre-established scoring tables allows NCDPI to take full advantage of short administration window for the mathematics NCEXTEND1 Alternate Assessments that are usually the last 5–10 days of the school year and is still able to provide and use scores for end of year reporting.

The final item statistics were based on the base year (2018–19) operational administration. There were two main rationales for using the operational administration statistics to develop raw to scale tables. First, the newly developed items for *Edition 4* were from 2017–18 standalone field-test where the new extended content standards may not have been fully implemented in class. Therefore, students may not have the opportunity to learn grade-level new extended content standards associated with the new items. As a result, item statistics for some new items, particularly those with revised or new content, were expected to be less reliable between field-test and operational administration. Relying on the observed score from the standalone field-test statistics under these circumstances for scoring would have resulted in larger-than-expected standard deviation of the raw scores as the sample size were also small resulting in the final raw-to-scale tables less stable.

The second rationale was that the new observed mean and standard deviation, used for scaling the raw to scale score, based on the student population from operational administration after students and schools offered opportunity to learn ensured a high degree of confidence of the scale. The classical summary statistics based on the operational 2018–19 administration are shown in *Table 6.1*.

Table 6.1 Average CTT Statistics, 2018–19 Operational Forms

Grade	N*	First Trial		Second Trial	
		P-Value	Point Biserial	P-Value	Point Biserial
3	~1274	0.46	0.39	0.74	0.41
4	~1205	0.41	0.36	0.72	0.40
5	~1230	0.44	0.34	0.71	0.34
6	~1241	0.39	0.35	0.72	0.39
7	~1264	0.42	0.33	0.72	0.32
8	~1177	0.44	0.36	0.73	0.38
10	~1047	0.45	0.31	0.73	0.37

*Sample size per item varied due to non-response on some items

When comparing the CTT statistics based on the 2017–18 standalone field test (*Table 4.1*) with 2018–19 operational administration (*Table 6.1*) a pattern emerged that the items on average are perceived easier across all grades in succeeding operational administration. The average form difficulty for the First Trial increased minimally, the P-Value difference was 0.03 or less from 2017-18 stand-alone field-test to 2018–19 operational administration. The difference, however, was larger in Second Trial with p-value differences ranging from 0.08 to 0.17. One of the assumptions, as indicated earlier, is that the field-test data may be inherently susceptible to opportunity to learn of the new extended content standards. The difference between the field-test and operational statistics emphasize the importance of evaluation of item performance in base year and use of student responses from operational administration for final base scale.

6.2 Scaling

Standard 5.2 (AERA, APA, & NCME, 2014) states that, “*the procedures for constructing scales used for reporting scores and the rationale for these procedures should be described clearly*” (p.102). Adhering to the standard NCDPI used linearly transformed scale score from raw score for the grades 3-8 and NC Math 1 mathematics NCEXTEND1 Alternate Assessment reporting. The transformation included descriptive statistics of raw scores (observed mean and standard deviation) from the 2018–19 operational administration. For the NCEXTEND scaling procedures the following linear transformation of raw scores were implemented:

$$Y_x = \left(\frac{9}{SD(x)} * RS \right) + \left(450 - \left(\frac{9}{SD(x)} * Mean(x) \right) \right)$$

Where Y_x is the scale score given a raw score (RS), $SD(x)$ is the standard deviation of observed RS, $Mean(x)$ is the observed mean of the RS from the 2018–19 operational administration. The expected theoretical mean of the scale scores is 450 and standard deviation is 9. During the base year (2018–19) of implementation of *Edition 4* Extended Content Standard, students’ scores

were delayed until after the standard setting workshop was completed and new performance achievement levels were adopted by the NCSBE.

6.3 Automated Decentralized Scoring

The NCEXTEND1 forms are administered electronically via a centrally hosted NCDPI-managed test server. The school system’s test coordinator downloads and imports student response data and runs the scoring software to generate scoring reports. Prior to the release of final results to schools, test coordinators perform quality control checks. They then provide results (reports) from the test administrations to their respective schools if no error was reported and after the NCDPI confirms its final score certification check was completed. Once the data are available, school system test coordinators can generate school rosters, class rosters, and individual reports. Initial district/school-level reporting occurs at the LEA level. North Carolina Administrative Code (i.e., *16 NCAC 06D .0302*) requires districts to report scores resulting from the administration of district-wide and state-mandated tests to students and parents or guardians along with available score interpretation information within 30 days from generation of the score at the district level or from the receipt of the score and interpretive documentation from the department.

6.4 Score Certification

Standard 6.9 (AERA, APA, & NCME, 2014) states, “*Those responsible for test scoring should establish and document quality control processes and criteria*” (p.118). Prior to the release of test scores for official reporting and use for further analyses, the NCDPI performs a final certification to ensure the correct answer key was used in all phases of the scoring to record students’ number correct scores. The certification process requires the completion of two main quality control steps: In the first step, the psychometric team using the recorded student response data independently tabulates the number correct score at the student level and compares that to the recorded number correct score reported by the scoring software. The goal is to have a 100% agreement rate between scores from the official scoring software and the independent check. The second step to complete the score certification process involves a sample review of CTT item statistics from operational forms. The goal is to check if current item level CTT statistics are consistent with the base year. During this step, if the form level statistics differed significantly it is further investigated at item level to make sure the scoring is correct. If any issues are found because of either a wrong scoring key or an improper rendering of any sort, the item is dropped from the form as an operational item and a new raw-to-scale table is generated for that form and the entire scoring procedure is updated with the new data.

Upon completion of score certification analyses, the generated test data are certified as accurate provided that all NCDPI-directed test administration guidelines, rules, procedures and policies have been followed at the district and school levels in conducting proper test administrations and in the generation of the student response data. Finally, the NCDPI issues an official communiqué affirming scores have been certified and scale scores are approved for official reporting.

CHAPTER 7 STANDARD SETTING

Standard setting is a process to define levels of achievement or proficiency and the cut scores corresponding to those levels. 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*” (p. 107). For the first operational administration of the *Edition 4* NCEXTEND1 2018–19 mathematics assessments, NCDPI contracted with the Data Recognition Corp DRC¹ to conduct a full standard setting workshop with the main goal of recommending achievement levels and cut scores for the newly developed assessments.

Since achievement levels or cut scores involve high-stakes decision-making including student, teacher and school level accountability, validity of the standard setting process and resulting cut scores is very important. Kane (2001) identified three elements of validity for standard setting: 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’ confidence in the cut score recommendations they made. Internal validity evidence suggests 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. The processes and evidences in abbreviated version of the *Edition 4* NCEXTEND1 mathematics final standard setting are presented in the ensuing sections. A full standard setting technical report produced by DRC can be found in [Appendix 7–A](#).

7.1 Standard Setting Activities

On July 8–11, 2019, the NCDPI and DRC conducted a standard setting for the NCEXTEND1 *Edition 4* mathematics tests in grades 3–8 and NC Math 1 Alternate Assessments. The purpose of the NCEXTEND1 standard setting was to develop achievement standards, achievement level descriptors (ALDs), and cut scores associated with three achievement levels: Not Proficient, Level 3 and Level 4. All together there were 37 participants for the standard setting of the NCEXTEND1 tests. Two panels; grades 3–6 with 20 participants and grades 6–8 and NC Math 1 with 17 participants of North Carolina math educators convened in Raleigh, North Carolina to

¹Copyright © 2019 Data Recognition Corp.

make cut score recommendations for the assessments. Both groups collaborated to recommend cut scores for the grade 6 with participants divided into seven (7) tables of approximately five (5) participants each. For subsequent grades, participants divided into their pre-assigned groups. Participants in each group were divided into four tables. One participant at each table served as the table leader. Table leaders moderated discussions at their tables and helped the workshop staff distribute and collect the secure workshop materials. The table leaders were not members of the workshop staff, and they contributed to their committees’ recommendations.

NCDPI used the Yes/No Angoff (Plake & Cizek, 2012) procedure to recommend cut scores for NCEXTEND1 mathematics. The Yes/No Angoff procedure has been used to establish achievement standards for educational assessments around the world. All training during the standard setting workshop was facilitated by the DRC staffs.

7.1.1 Panelists’ Backgrounds

Table 7.1 shows the gender and ethnicity distributions of the workshop participants. As demonstrated by the information provided in these tables, panelists making up the standard-setting panels showed representation of diverse gender and ethnic background across North Carolina with majority of the participants as female (92%) and white (78%).

Table 7.1 Panelist Gender and Ethnicity

N	Gender			Ethnicity						
	F	M	NR	AA	AI	HI	NA	WH	MI	NR
37	92%	5%	3%	13%	3%	0	0	78%	3	3

F=Female, M=Male, NR=No Response, AA=Black, AI=American Indian, HI=Hispanic, NA=Not Applicable, WH=White, and MI=Mixed

All panelists were asked to provide voluntary experience information. A brief summary of panelists’ experiences in terms of years in current position and professional background are presented in **Error! Reference source not found.** through **Error! Reference source not found.** **Error! Reference source not found.** illustrates the educational experience of the panelists in terms of the years in current position. It shows the experience ranged from five (5) years or less to twenty-five (25) years or more indicating a very diverse group of educators participated in the standard setting workshop.

Table 7.2 Panelist Experience as Educators (%)

N	Years in Current Position						
	1–5	6–10	11–15	16–20	21–25	Over 25	No Response
37	5%	24%	16%	19%	16%	16%	3%

The panelists’ professional backgrounds in terms of teaching diverse group of students are summarized in **Error! Reference source not found.** and **Error! Reference source not found.**. These tables show that the teachers from diverse experience including teaching general education, special education, ELs, gifted and talented as well as higher education involved in the standard setting.

Table 7.3 Panelist Professional Background: Three–Grade Panels

N	General Education Teacher	Special Education Teacher	ELL Teachers	Curriculum Staff	Higher Education	Teachers on Special Assignment	Administrator	No Response
37	54%	19%	3%	19%	0%	0%	3%	3%

Table 7.4 Panelist Professional Background: Single–Grade Panels

N	Special ed. in a self-contained classroom	Special ed. in a mainstream classroom	English learners	Gifted and talented ed.	Vocational ed.	Alternate ed.	Adult ed.	No Response
37	22%	73%	51%	41%	11%	3%	5%	8%

7.1.2 Opening Session and Introductions

All participants began the workshop with a single opening session for the general and NCEXTEND1 Alternate Assessments led by the NCDPI. During this session, director of the NCDPI Accountability Division welcomed the participants to the workshop and described the purpose of the workshop. Subsequently, the section chief of Test Development described the recent changes to the tests, and how valuable the participating educators’ recommendations would be in identifying new cut scores for the tests.

Following the committee introductions, the two-grade level panels (grades 3–5, grades 6–8 and NC Math 1) for the NCEXTEND1 standard setting spent the remainder of Monday, July 8, discussing achievement level descriptors (ALDs) drafted by the NCDPI in consultations with state educators. The ALDs serve as content-oriented statements describing expectations of student performance at each achievement level. Breakout-session facilitators provided panelist with ALD training that covered the purpose of ALDs, and facilitators shared several real-world examples demonstrating characteristics of effective ALDs. Panelists were trained on strategies to link ALDs to the test blueprint and curriculum standards, both of which were made available to panelists. The NCDPI provided policy ALDs for the general mathematics tests in advance of the standard setting workshop, which included general and policy-oriented statements about student achievement across levels. Panelists were tasked with adding content-oriented statements to the

draft ALDs to further define student achievement in the context of the assessment. The panels’ final drafted ALDs were turned over to the NCDPI for review and future revisions, as deemed necessary.

7.1.3 Achievement Level Descriptors

The ALDs summarize the knowledge, skills, and abilities expected of students in each achievement level. Three types of ALDs summarized here included policy ALDs, Range ALDs, and Threshold ALDs. The NCDPI pre worked on ALDs development process by drafting the initial ALDs, rounds of webinars, and revisions with the North Carolina educators to finalize. The descriptions of Not Proficient or Inconsistent Understanding, Level 3 or Sufficient Understanding, and Level 4 or Thorough Understanding are policy ALDs (*Table 7.5*) for public statements about what and how much North Carolina educators want students to know and be able to do for each grade level in NCEXTEND1 Mathematics. Level 4 students are also considered on track for competitive employment and post-secondary education.

Table 7.5 Policy ALDs for NCEXTEND1 Alternate Mathematics Assessment

Not Proficient	Level 3	Level 4
Students at the Not Proficient level demonstrate inconsistent understanding of grade level content standards and will need support at the next grade/course.	Students at Level 3 demonstrate sufficient understanding of grade level content standards though some support may be needed to engage with content at the next grade/course.	Students at Level 4 demonstrate a thorough understanding of the North Carolina Extended Content Standards and are on track for competitive employment and postsecondary education.

Range ALDs summarize the knowledge, skills, and abilities expected of students in a given achievement level on a specific test. The range ALDs show the types of content, as informed by the state content standards, that should be mastered by students in each achievement level on the test at hand. Threshold ALDs are based on the range ALDs and summarize the knowledge, skills, and abilities expected of students who are at the point-of-entry (the threshold) of each achievement level. For any given test, these descriptors show the types of skills needed just to be classified in a given achievement level (e.g., just to be classified in Level 3). NCDPI provided policy ALDs for the NCEXTEND1 mathematics tests in advance of the standard setting workshop. At the standard setting, participants worked to develop formal range ALDs (on Day 1) and informal threshold ALDs (on Days 2–4). The range ALDs are shown in Section E of the Standard Setting Technical Report (*Appendix 7–A*).

7.1.4 Method and Procedure

The Angoff (1971) procedure is one of the most implemented methods to establish achievement standards on educational assessments. In one modification, panelists review each item and estimate what proportion of a hypothetical group of hypothetical threshold examinees would answer each item correctly (Zieky, 2012). Several modifications to this original procedure have been implemented. The Yes/No Angoff method addresses two difficulties that panelists may have in applying the procedure (Impara & Plake, 1997). First, panelists may have difficulty in conceptualizing the hypothetical threshold students. Second, estimating the proportion correct may be a difficult task even for a clearly defined group of examinees. In the Yes/No method, panelists are directed to make a dichotomous (“Yes” or “No”) judgment about whether the hypothetical threshold examinees would be able to answer each question correctly.

The Yes/No Angoff (Plake & Cizek, 2012) method is well-suited to assessments comprised entirely (or predominantly) of selected-response items, like the NCEXTEND1. Also, the Yes/No Angoff method was selected over other standard setting procedures, notably item-mapping procedures like the Bookmark Standard Setting Procedure (Lewis, Mitzel, & Green, 1996) because of the smaller sample size of the NCEXTEND1 Alternate Assessment testing population.

7.1.5 Across-Grade Articulation and Final ALD Cuts

Throughout the standard setting process, participants were informed they would have an opportunity at the end of the workshop to consider the across-grade articulation of the achievement standards. Participants were told that achievement standards were well articulated when the impact data associated with a set of cut scores formed a reasonable, explainable pattern across grades.

During the across-grade articulation, table leaders were assembled in a room and DRC examined the ranges of cut score recommendations made by participants during the standard setting. The table leaders were briefed that cut scores adopted within the ranges are considered as reflecting the voice of the standard setting committee. Subsequently, DRC presented the adjusted cut scores and associated impact data to the table leaders for their inspection. The group saw how the adjustments reflected their opinions about the articulation of the students in Not Proficient and in Level 3 and above. DRC asked the group whether it felt comfortable making this set of adjusted cut scores its recommendation and the table leaders assented. DRC reminded the table leaders that NCDPI and its advisors would be reviewing their cut score recommendations and that adjustments may be made to the cut scores by NCDPI for policy-related reasons.

After the revision, the final achievement level Cuts (*Table 7.6*) were presented to the NCSBE on August 7, 2019 for consideration. After deliberation, the NCSBE approved the cut scores on August 8, 2019. The 2018-19 NCEXTEND1 population raw score ranges for the achievement level cuts are shown in *Table 7.7*.

Table 7.6 Final NCEXTEND1 Recommended Cuts and Proficiency Distributions

Grade	Recommended Cuts		Percent (%) of Students in Each Achievement Level Based on Recommended Cut Scores		
	Level 3	Level 4	Not Proficient	Level 3	Level 4
3	451	464	56.0	37.4	6.7
4	451	465	58.6	35.2	6.2
5	452	465	63.6	28.8	7.6
6	453	464	58.4	36.1	5.6
7	450	467	52.5	42.4	5.1
8	453	465	67.2	26.0	6.9
10	452	463	56.8	37.8	5.5

Table 7.7 2018-19 Population NCEXTEND1 Raw Score Ranges Across Achievement Levels

Grade	Not Proficient		Level 3		Level 4	
	Min	Max	Min	Max	Min	Max
3	0	33	34	46	47	54
4	0	31	32	44	45	54
5	0	32	33	44	45	53
6	0	32	33	43	44	53
7	0	30	31	44	45	54
8	0	34	35	45	46	54
10	0	34	35	44	45	53

7.2 Evaluation of the Standard Setting Workshop

Since standard setting process incorporates subjective expert judgement, it is important to document procedural validation including selection of the experts, experts’ clarity of the standard setting method and their judgement, i.e., the extent to which they understand the standard setting procedure and their confidence in the cut scores. Sections below summarizes the participants’ evaluation of the process as well as evaluation of the processes by the external evaluator.

7.2.1 Participants' Evaluation

At the end of the workshop, a participant survey was conducted for their perceived validity of the workshop and their recommendations as a part of the post-session workshop evaluation. Such evaluations are important evidence for establishing the validity of performance levels (Hambleton, 2001). The survey results are presented in *Table 7.8*. Generally, 97% or higher proportion of participants were satisfied (Agree or Strongly Agree) with their recommendations and with the workshop. The results further indicated that 100% of the participants understood and considered the threshold students when making benchmarks. They agreed that the final recommended cut scores reflected the work of their group.

Table 7.8 NCEXTEND1 Standard Setting Workshop Evaluation Results

Statement	Strongly Disagree	Disagree	Agree	Strongly Agree	Agree + Strongly Agree
The training provided a clear description of the workshop goals.	0%	3%	78%	19%	97%
I understood how to make my bookmarks.	0%	0%	75%	25%	100%
I considered the threshold students when making my bookmarks.	0%	0%	36%	64%	100%
Discussing the threshold students helped me make my bookmarks.	0%	3%	39%	58%	97%
My group's work was reflected in the presentation of recommendations across grades.	0%	3%	64%	33%	97%
Overall, I valued the workshop as a professional development experience.	0%	3%	61%	36%	97%

7.2.2 External Evaluation

In order to implement and evaluate any deviations from the standard setting processes by the vendor, the NCDPI contracted Dr. Gregory J. Cizek as an external independent evaluator of the mathematics standard setting workshop. Dr. Cizek is an expert in the field and is also a member of the North Carolina Technical Advisory Committee (NCTAC). His evaluation report regarding the standard setting workshop in general and process in particular are summarized below. The observation report is available in [Appendix 7–B](#).

Dr. Cizek reported that qualified educators from North Carolina were trained in the methods and led through the standard setting procedures by content and process specialists. Dr. Cizek concluded that “the workshop recommended cut score can be considered to be valid and reliable estimates of appropriate performance standards for the relevant assessments. Unless the panelists’ evaluations indicate otherwise, policy makers should have confidence that the recommendations from the standard setting activity are based on sound procedures, producing credible, defensible, and educationally useful results.”

CHAPTER 8 TEST RESULTS AND REPORTS

This chapter presents test level summary results for the alternate mathematics assessment based on reported scale scores and achievement levels from 2018–19 NCEXTEND1 mathematics operational administration. The chapter is divided into three main sections. Section 8.1 highlights descriptive summary results of scale scores overall and by major demographic subgroups including accommodations, gender, and ethnicity as well as overall achievement level distributions for the NCEXTEND1 Alternate Assessments. Section 8.2 briefly describes types of reports NCDPI produces including those at class, school, district, and state level to share and interpret assessments results with stakeholders. Section 8.3 elaborates confidentiality requirements for sharing or reporting students’ personal information as well as student data.

8.1 NCEXTEND1 Scale Score Distribution

Scale score distributions for the *Edition 4* mathematics grades 3-8 and NC Math 1 NCEXTEND1 results from 2018–19 operational administration are summarized in *Figure 8.1* through *Figure 8.8*. The results are based on all eligible students enrolled at the grade level NCEXTEND1 Alternate Assessments.

The population scale score mean for the NCEXTEND1 raw to scale scores linear transformation was set to 450 with a standard deviation of 9. Note that *Edition 4* NCEXTEND1 mathematics scale scores across grades are not in vertical scale. Any across-grade scale score interpretations and comparisons are not recommended as each NCEXTEND1 assessment is aligned to grade level specific content standards.

Results show that scale score distributions from NCEXTEND1 administration have similar distributional properties as the scaling parameters with about a mean of about 450 and standard deviation of nine.

Figure 8.1 Grade 3 NCEXTEND1 Mathematics Scale Score Distribution, Spring 2019

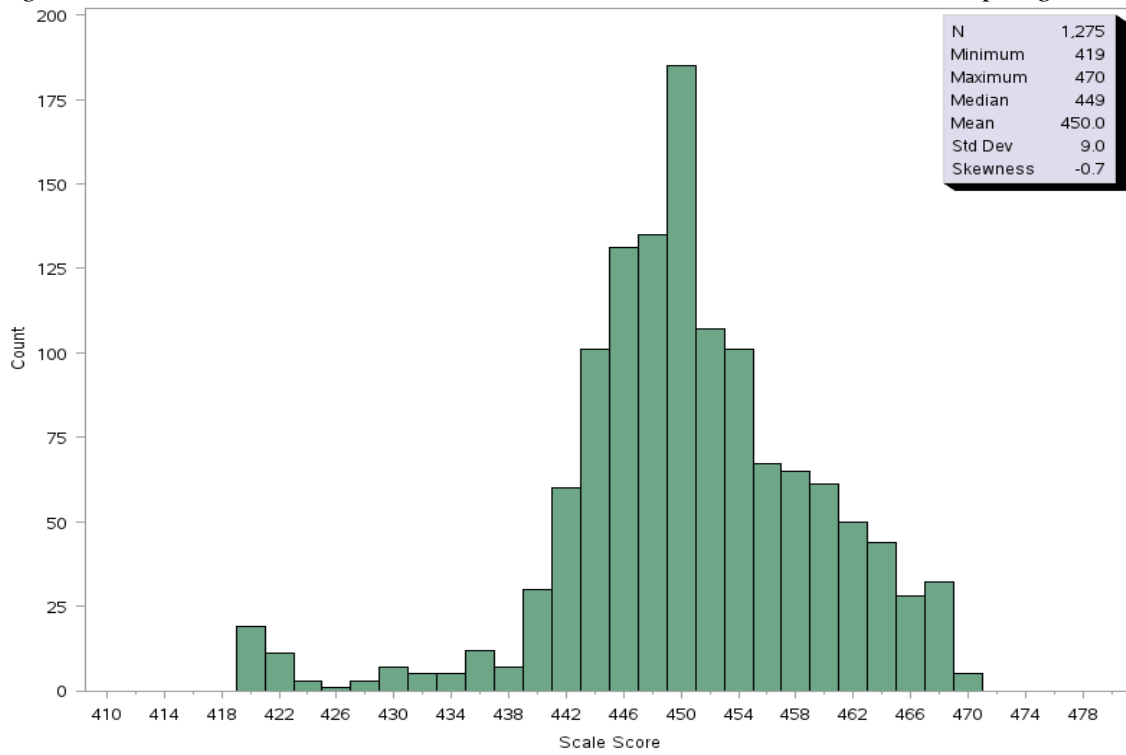


Figure 8.2 Grade 4 NCEXTEND1 Mathematics Scale Score Distribution, Spring 2019

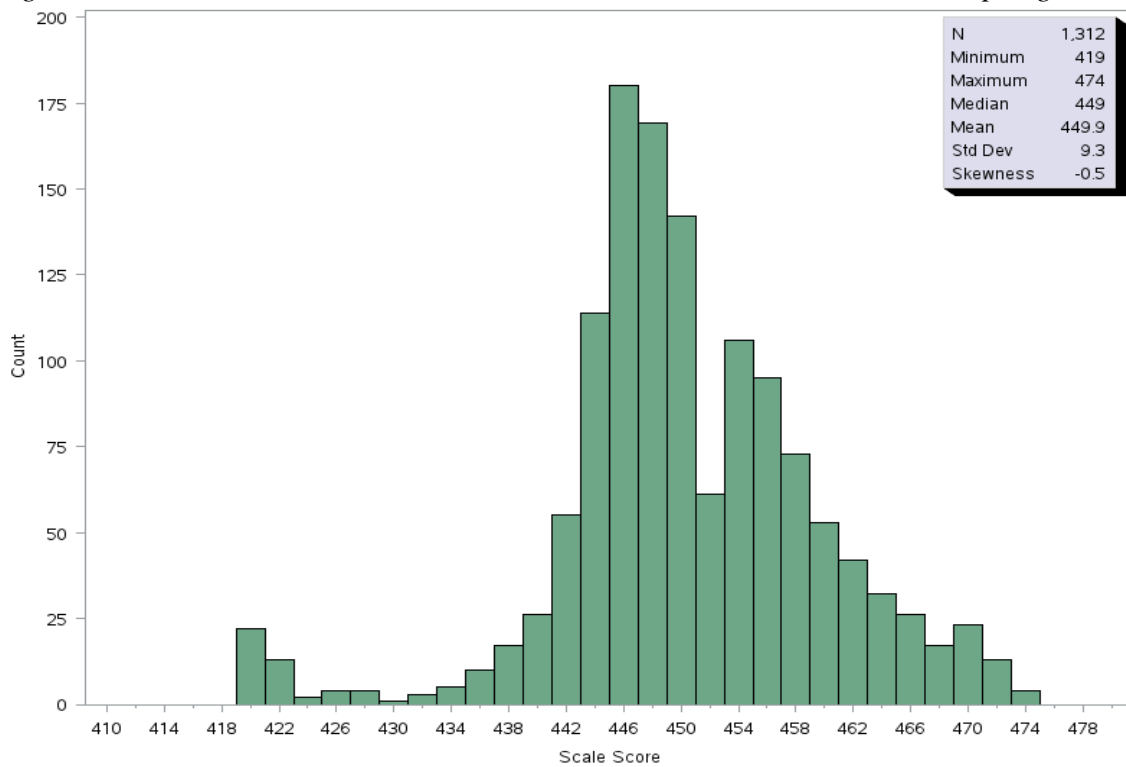


Figure 8.3 Grade 5 NCEXTEND1 Mathematics Scale Score Distribution, Spring 2019

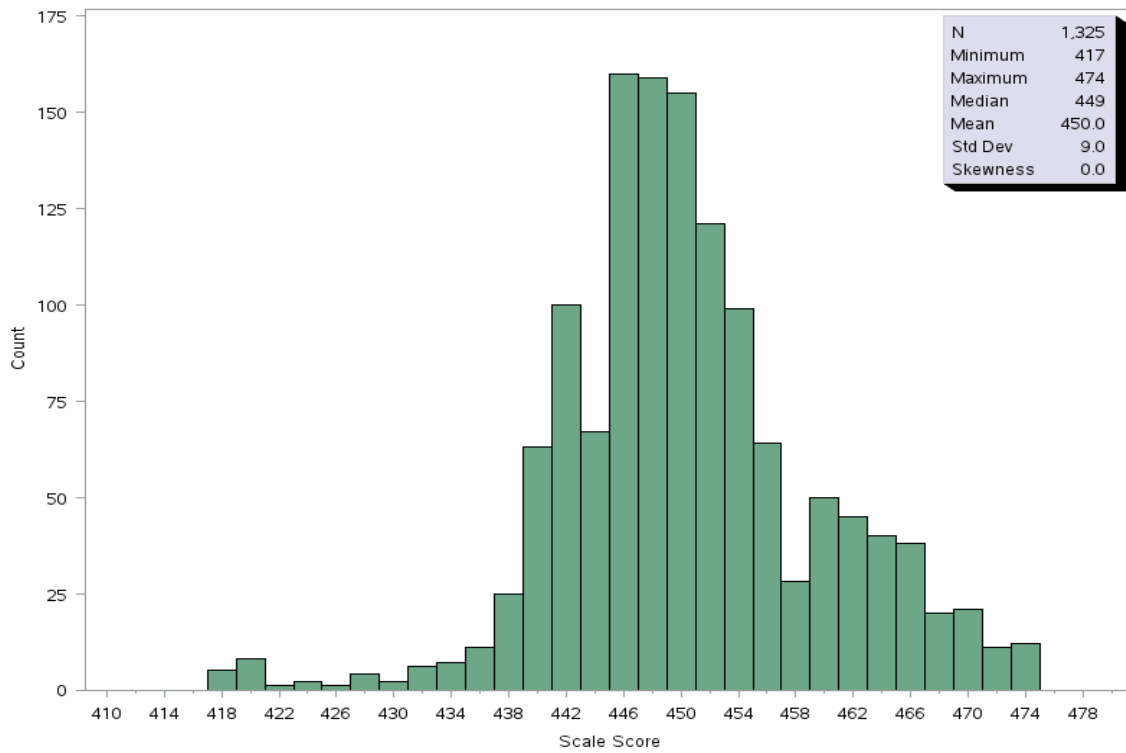


Figure 8.4 Grade 6 NCEXTEND1 Mathematics Scale Score Distribution, Spring 2019

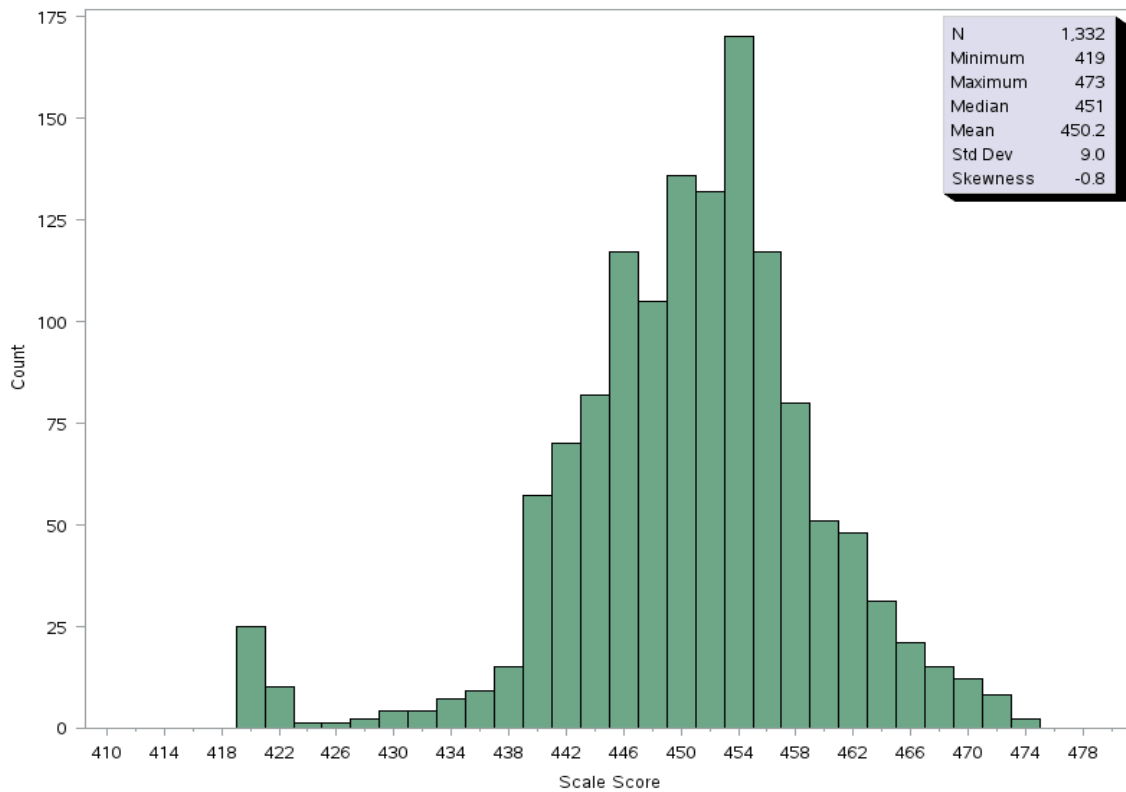


Figure 8.5 Grade 7 NCEXTEND1 Mathematics Scale Score Distribution, Spring 2019

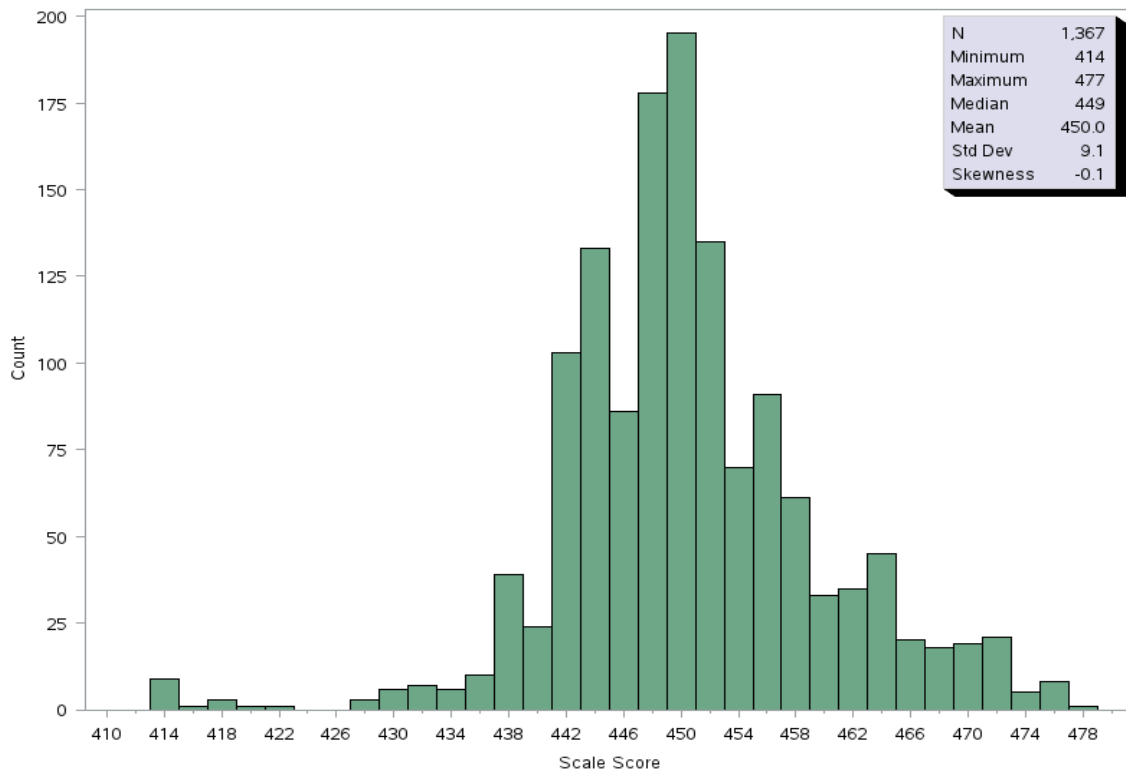


Figure 8.6 Grade 8 NCEXTEND1 Mathematics Scale Score Distribution, Spring 2019

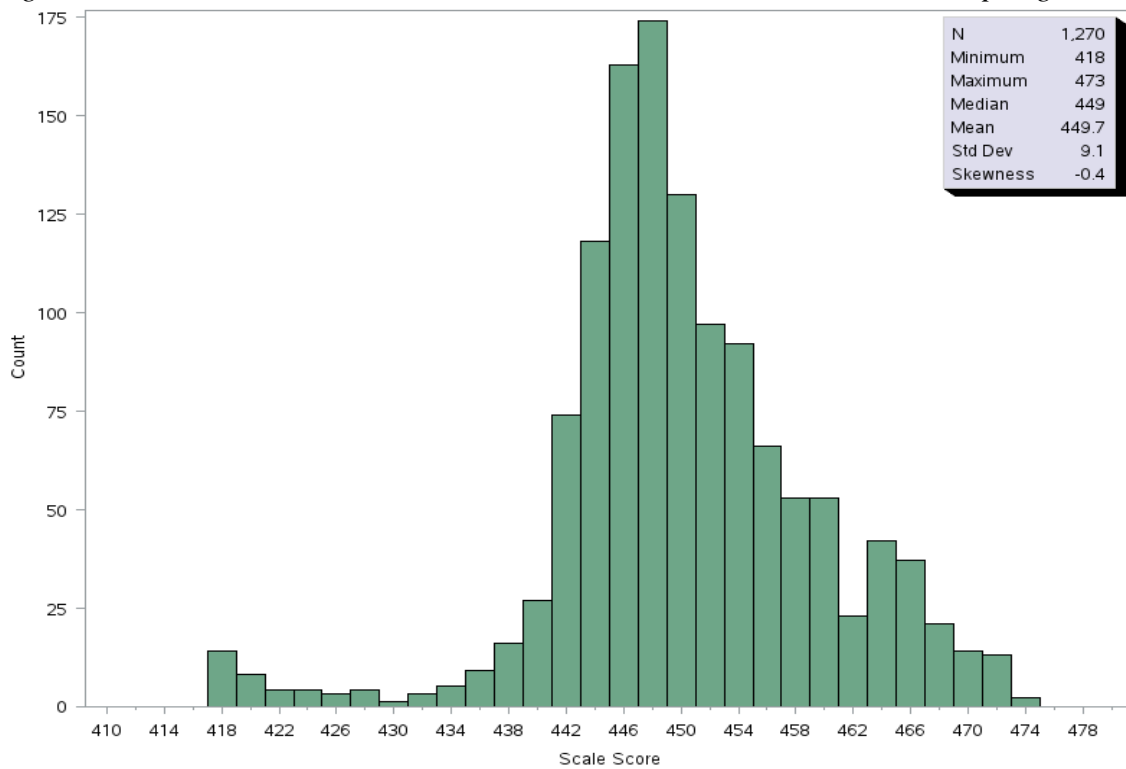
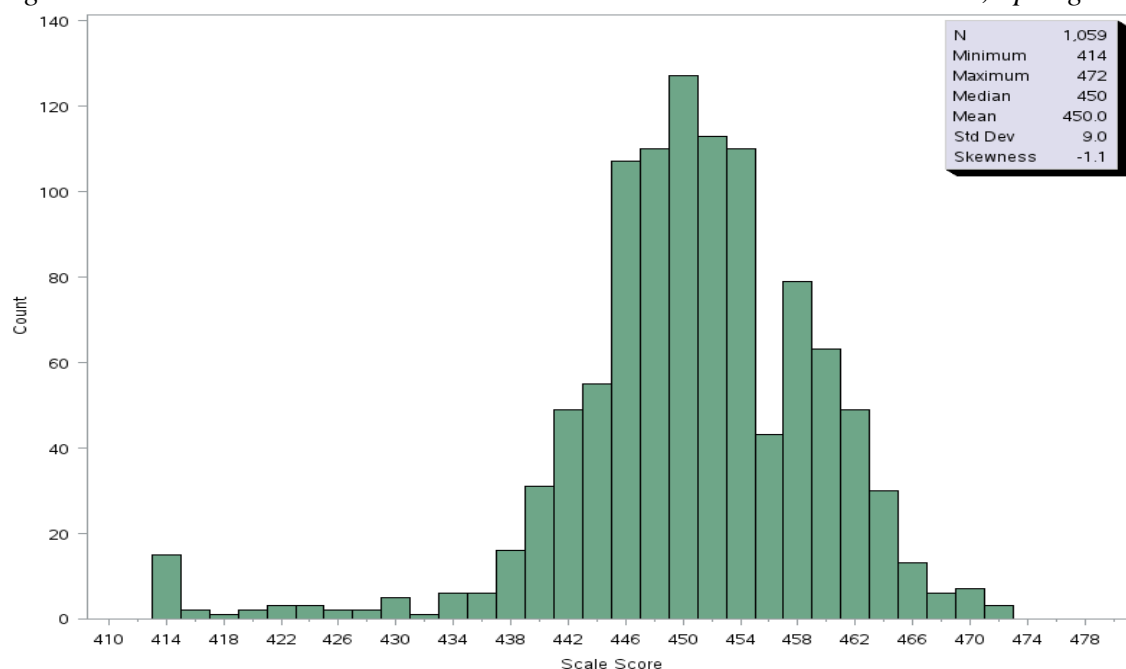


Figure 8.7 NC Math1 NCEXTEND1 Mathematics Scale Score Distribution, Spring 2019



8.1.1 Scale Scores by Disability Subgroups

The NCDPI allows the use of various types of accommodations in NCEXTEND1 assessments to ensure accessibility to all students. Students with IEPs can access their required accommodations described in Chapter 5 at any time during test administration. Research in measurement literature has demonstrated that these standard accommodations do not measure any significant construct irrelevant variance to students reported scores. Thus, students’ scores from these approved accommodations are included in the NCEXTEND1 overall results and the same inferences as the regular NCEXTEND1 are made about student’s performance. The scale score descriptive summary results for grades 3–5 are shown in *Table 8.1*, and grades 6–8 and NC Math 1 are shown in *Table 8.2*. The major accommodation subgroups containing approximately 10% or more students are summarized separately. Those less than 10% are combined and label as “Other”.

These results show that scale score distributions from NCEXTEND1 administration have similar distributional properties as the scaling parameters with about a mean of 450 and standard deviation of nine. For all grades, Autism accommodation was the most used accommodation category followed by Moderate Intellectual Disability, Mild Intellectual Disability, and Multiple Disability. The average scale score for Mild Intellectual Disability was the highest across all grades.

Table 8.1 Grades 3–5 NCEXTEND1 Mathematics Scale Score by Accommodation Subgroups, Spring 2019

Grade	Students with Disability Category	N	Statistics		Range		Percentile		
			Average	SD	Min	Max	25th	Median	75th
3	Autism	539	451	9	419	469	446	450	457
	Intellectual Disability - Mild	169	454	7	431	470	449	454	459
	Intellectual Disability - Moderate	256	449	7	421	468	445	449	453
	Multiple Disability	163	445	10	419	468	442	446	451
	Other	148	449	10	419	468	445	449	455
	All	1,275	450	9	419	470	445	449	455
4	Autism	535	451	9	419	474	446	449	456
	Intellectual Disability - Mild	179	453	8	423	471	447	453	458
	Intellectual Disability - Moderate	276	449	8	419	473	445	448	453
	Multiple Disability	154	447	11	419	471	443	448	453
	Other	168	449	11	419	474	445	448	455
	All	1,312	450	9	419	474	445	449	455
5	Autism	493	450	9	419	474	445	449	455
	Intellectual Disability - Mild	192	454	9	434	473	448	452	460
	Intellectual Disability - Moderate	309	449	8	419	474	445	448	453
	Multiple Disability	142	447	10	417	471	442	447	452
	Other	189	450	10	417	474	444	448	456
	All	1,325	450	9	417	474	445	449	454

Table 8.2 Grades 6–8 and NC Math 1 NCEXTEND1 Mathematics Scale Score by Accommodation Subgroups, Spring 2019

Grade	Students with Disability Category	N	Statistics		Range		Percentile		
			Average	SD	Min	Max	25th	Median	75th
6	Autism	478	451	9	419	472	445	452	456
	Intellectual Disability - Mild	196	454	6	438	470	449	454	457
	Intellectual Disability - Moderate	307	451	8	419	473	445	450	455
	Multiple Disability	166	445	11	419	469	441	447	451
	Other	185	450	9	419	472	445	450	455
	All	1,332	450	9	419	473	445	451	456
7	Autism	466	451	9	414	477	445	449	455
	Intellectual Disability - Mild	220	453	9	433	476	447	452	460
	Intellectual Disability - Moderate	348	449	8	414	471	444	449	454
	Multiple Disability	173	446	10	414	471	443	447	451
	Other	160	450	9	414	474	444	449	455
	All	1,367	450	9	414	477	444	449	455
8	Autism	426	450	9	418	472	445	448	455
	Intellectual Disability - Mild	194	455	8	428	472	450	455	461
	Intellectual Disability - Moderate	333	449	7	418	473	445	448	452
	Multiple Disability	157	445	10	418	469	442	446	450
	Other	160	449	10	418	473	444	448	454
	All	1,270	450	9	418	473	445	449	455
10	Autism	354	450	9	414	470	445	449	455
	Intellectual Disability - Mild	146	455	6	434	471	451	454	459
	Intellectual Disability - Moderate	286	451	7	416	472	447	450	455
	Multiple Disability	139	446	11	414	467	441	447	452
	Other	134	449	11	414	471	445	451	457
	All	1,059	450	9	414	472	446	450	455

8.1.2 Scale Scores by Gender

Table 8.3 shows scale score descriptive summary statistics for the mathematics grades 3–5 NCEXTEND1 tests by gender and *Table 8.4* shows for grades 6–8 and NC Math 1. Across grade levels, there were higher proportion of male students (66%–69%) who took NCEXTEND1

mathematics tests in North Carolina during 2018–19 school year. Across grades, male students on average performed one scale score point higher than female in most grades except grades 6 and NC Math 1 where the mean scale scores for both male and female were the same.

Table 8.3 Grades 3–5 NCEXTEND1 Mathematics Scale Score by Gender, Spring 2019

Grade	Gender	N	Statistics		Range		Percentile		
			Average	SD	Min	Max	25 th	Median	75 th
3	Female	400	449	9	419	468	445	449	454
	Male	875	451	9	419	470	446	450	457
	All	1,275	450	9	419	470	445	449	455
4	Female	417	449	9	419	473	445	448	454
	Male	895	451	9	419	474	445	449	456
	All	1,312	450	9	419	474	445	449	455
5	Female	421	449	9	417	474	444	448	453
	Male	904	451	9	417	474	445	449	455
	All	1,325	450	9	417	474	445	449	454

Table 8.4 Grades 6–8 and NC Math 1 NCEXTEND1 Mathematics Scale Score by Gender, Spring 2019

Grade	Gender	N	Statistics		Range		Percentile		
			Average	SD	Min	Max	25 th	Median	75 th
6	Female	444	450	9	419	473	445	450	455
	Male	888	450	9	419	473	445	451	456
	All	1,332	450	9	419	473	445	451	456
7	Female	460	449	9	414	476	444	449	455
	Male	907	450	9	414	477	445	449	455
	All	1,367	450	9	414	477	444	449	455
8	Female	423	449	9	418	472	445	448	453
	Male	847	450	9	418	473	445	449	455
	All	1,270	450	9	418	473	445	449	455
NC Math 1	Female	353	450	9	414	471	446	450	455
	Male	706	450	9	414	472	446	450	455
	All	1,059	450	9	414	472	446	450	455

8.1.3 Scale Score by Major Ethnic Groups

For the purpose of this report, the scale scores descriptive statistics for the NCEXTEND1 Alternate Assessments are summarized only for students who self-reported to be belonged to one of the major ethnic groups: Black, Hispanic, and White. Students not self-identified in any of those three major groups are classified as Other. *Table 8.5* and *Table 8.6* show the breakdown of NCEXTEND1 mathematics scale scores by major reportable ethnic groups from 2018–19 administration. The distribution of North Carolina alternate student population is very similar across grade levels with White students representing about 36% – 46% of all students across all grades and Black students representing about 33% – 36% with Hispanic students making about 12% – 19%. The average scale scores within a grade across ethnic groups are either the same or a maximum difference of two (2) scale score points.

The scale score differences represented in *Table 8.5* and *Table 8.6* are not necessarily an indication that the NCEXTEND1 assessments are biased across ethnic groups. All NCEXTEND1 items were thoroughly vetted throughout several phases of item development, field test and item analysis by different experts to ensure operational NCEXTEND1 mathematics items did not exhibit sensitivity to any student subgroup. The descriptive statistics of the scale scores for subgroups (Ethnicity, SWD, EDS, and ELs in extended form) are shown in [Appendix 8–A](#).

Table 8.5 Grades 3–4 NCEXTEND1 Mathematics Scale Score by Ethnicity, Spring 2019

Grade	Ethnic	N	Statistics		Range		Percentile		
			Average	SD	Min	Max	25th	Median	75th
3	Black	427	451	9	419	469	446	450	456
	Hispanic	240	450	8	419	470	445	449	455
	Other	128	450	9	419	469	445	449	455
	White	480	450	9	419	469	445	450	456
	All	1,275	450	9	419	470	445	449	455
4	Black	464	450	9	419	473	445	449	455
	Hispanic	242	451	9	419	474	446	450	457
	Other	137	450	11	419	472	445	450	456
	White	469	450	9	419	472	445	449	455
	All	1,312	450	9	419	474	445	449	455
5	Black	468	451	9	417	474	445	450	456
	Hispanic	228	450	8	417	473	445	449	453
	Other	111	449	9	428	471	442	448	453
	White	518	450	9	418	474	445	449	454
	All	1,325	450	9	417	474	445	449	454

Table 8.6 Grades 6–8 and NC Math 1 NCEXTEND1 Mathematics Scale Score by Ethnicity, Spring 2019

Grade	Ethnic	N	Statistics		Range		Percentile		
			Average	SD	Min	Max	25th	Median	75th
6	Black	473	451	9	419	473	446	451	456
	Hispanic	205	450	8	419	470	445	450	455
	Other	131	449	11	419	473	444	450	456
	White	523	450	9	419	472	445	451	456
	All	1,332	450	9	419	473	445	451	456
7	Black	462	450	9	414	476	444	449	455
	Hispanic	211	450	8	429	471	444	449	454
	Other	102	451	11	414	475	445	450	457
	White	592	450	9	414	477	444	449	455
	All	1,367	450	9	414	477	444	449	455
8	Black	440	450	9	418	473	445	449	455
	Hispanic	193	449	9	418	471	444	448	454
	Other	90	450	9	420	471	445	450	457
	White	547	449	9	418	473	445	449	454
	All	1,270	450	9	418	473	445	449	455
10	Black	352	451	8	414	470	446	450	455
	Hispanic	130	449	9	414	469	446	451	454
	Other	90	451	9	418	467	446	451	457
	White	487	450	10	414	472	445	450	457
	All	1,059	450	9	414	472	446	450	455

8.1.4 Achievement Levels Distributions

Beginning in 2018–19 with *Edition 4* NCEXTEND1 tests, the NCDPI transitioned to classify and report student performance on the grades 3–8 and NC Math 1 NCEXTEND1 mathematics using three (3) performance or achievement levels aligned to grade level content standards and policy expectations. The three achievement levels presented in Chapter 7 are:

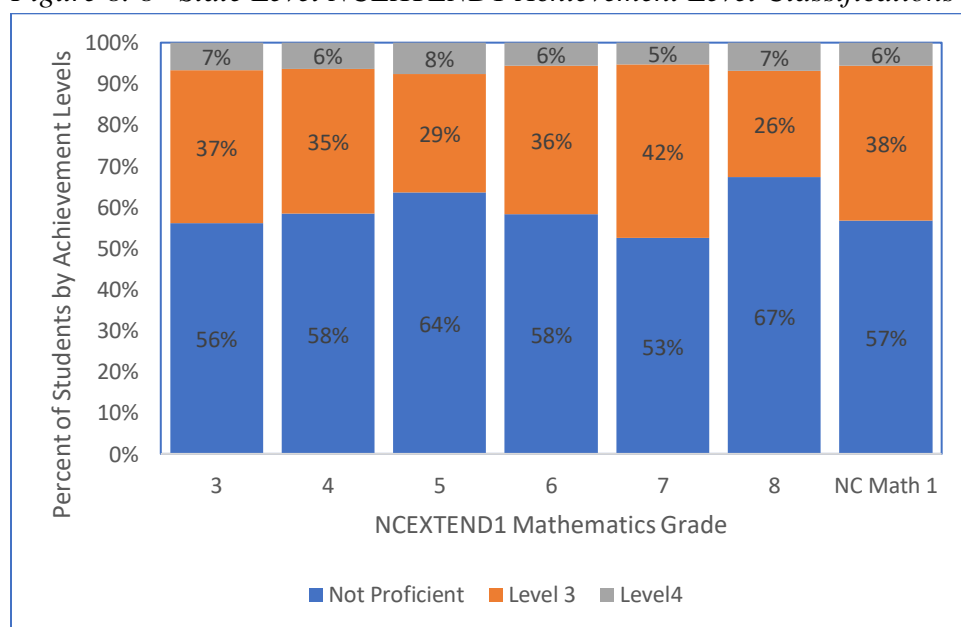
- **Not Proficient:** Students demonstrate inconsistent understanding of grade level extended content standards and will need support at the next grade/course.
- **Level 3:** Students demonstrate sufficient understanding of grade level extended content standards though some support may be needed to engage with content at the next grade/course.
- **Level 4:** Students demonstrate a thorough understanding of grade level extended content standards and are on track for competitive employment and postsecondary education.

These policy descriptors are used to summarize performance expectations for students at each level. For a detailed explanation of what students in each performance level are expected to be able to do refer to the full achievement level descriptors in [Appendix 8–B](#). These achievement levels with their associated achievement level descriptors represent the principal standards-based claims that the NCDPI has sufficient validity evidence for interpreting students’ NCEXTEND1 mathematics scores.

Based on the NC state law prescribed in the state accountability model, all students with NCEXTEND1 performance levels of Level 3 and Level 4 are considered and reported to have met grade level performance expectations. Students classified as Level 4 are further designated to be on track for competitive employment and postsecondary education. The Level 4 students are also used for federal accountability.

Figure 8.8 shows summary of proportion of students by achievement level classifications from the 2018–19 NCEXTEND1 mathematics assessments. The stacked bar graph shows classifications by grade. For example, 56% students in grade 3 are classified as Not Proficient, 37% Level 3, and 7% Level 4 indicating 44% (Level 3 and above) of NC grade 3 students who took the NCEXTEND1 mathematics assessment are considered to have met grade level content expectations for state accountability reporting purposes. While about 7% of these students are considered proficient and on-track for competitive employment and postsecondary education. The achievement level classifications for subgroups (gender, ethnicity, SWD, EDS, and ELs in extended form) are shown in [Appendix 8–C](#).

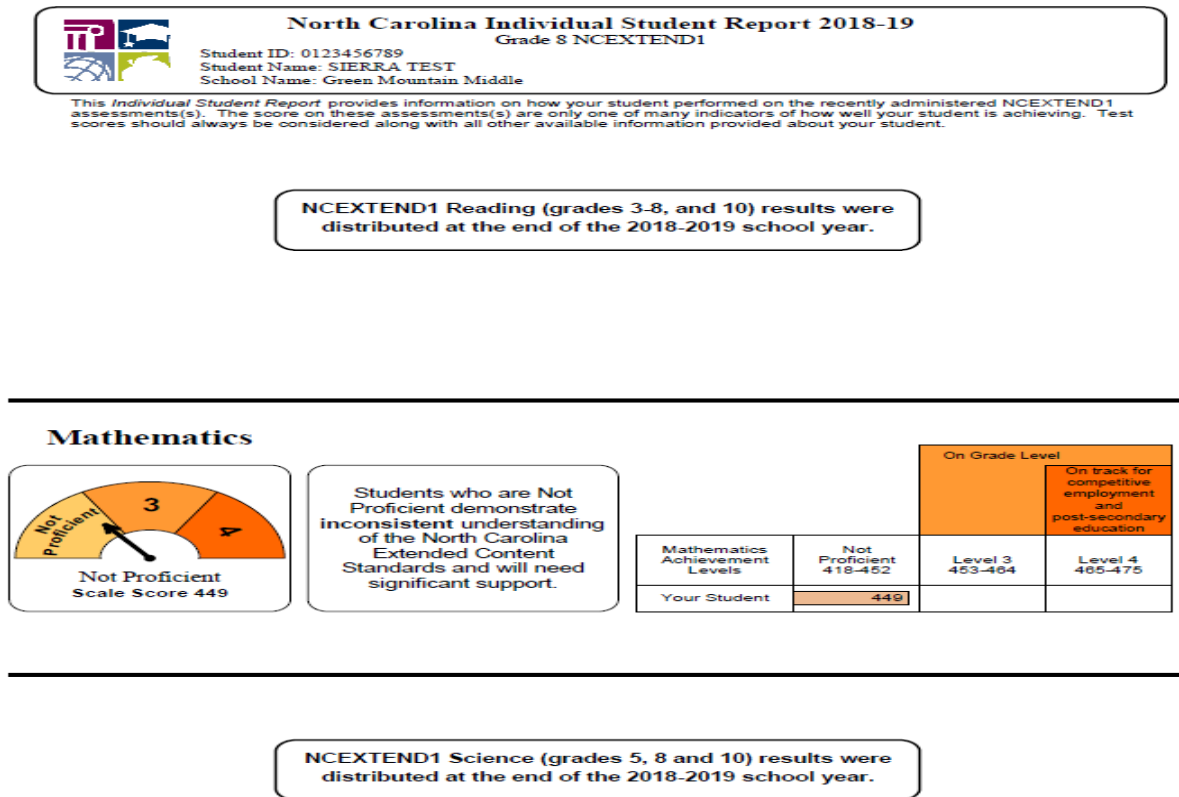
Figure 8. 8 State Level NCEXTEND1 Achievement Level Classifications by Grade, Spring 2019



8.2 Score Reports

Consistent with Standard 1.1 (AERA, APA, & NCME, 2014) which states, “*Test developers should set forth clearly how test scores are intended to be interpreted and consequently used*” (p. 23), annual results from NCEXTEND1 assessments are compiled and reported in a variety of formats for two main audiences. The first audience reporting category is for individual students and their parents/guardians. The individual reports for the NCEXTEND1 population present much of the same information as the ISRs for the general assessment. The Individual Student Report (ISR) example shown in *Figure 8.9* is designed to inform students, teachers, students’ parents, and school administrators on their overall performance based on the assessment as it relates to their standing on grade level content. The ISR highlights the achievement level and descriptor, with the associated scale score, the student is classified into based on performance. More information and description of the ISR is available on the NCDPI website. <http://www.dpi.state.nc.us/accountability/policies/uirs>.

Figure 8.9 Individual Student Report (ISR)



For more information on Individual Student Reports, please visit <http://www.dpi.state.nc.us/accountability/policies/uirs>

The second set of reports are generally generated for school and district audiences aimed to provide teachers and school administrators with in-depth and disaggregated data of their students and school performance to help inform instructional policies. In the current report format these reports are available as flat files that are pre-programmed in the reporting system and distributed to schools upon request. The goal, moving forward, is to have these reports in query database format so schools and district, will be able to run custom report, in real time. *Table 8.7* shows a summary list of the main pre-programmed static reports that are currently available to the different audiences for NCEXTEND1 mathematics assessments. The NCDPI also publishes, on its website, interpretive guides intended to help educators and decision makers at the classroom, school, and district levels understand the content and uses of the various score reports (See [Appendix 8–D](#)). These guides are also intended to help administrators and educators explain test results to parents and to the public.

Table 8.7 Reports by Audience

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

8.3 Confidentiality of Student Information

Regarding the confidentiality of student information, the State Board of Education policy *GCS-A-010 (j)(1)* clearly states that “*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.*” Sections below briefly describe the NCDPI’s guidelines for confidentially handling students’ personal information and test data.

8.3.1 Confidentiality of Personal Information

The *North Carolina Test Coordinators' Policies and Procedures Handbook* instructs that while handling and transmitting personally identifiable information, employees of Public School Units (PSU) of the NCDPI or other education institutions are legally and ethically obliged to safeguard the confidentiality of any private information they access while performing official duties. To protect the confidentiality of individuals from those who are not authorized to access individual-level data, Personally Identifiable Information (PII) is encrypted during transmission using one of the following methods, in order of preference:

- Secure FTP Server based on SFTP or FTPS protocols - Preferred method and most widely acceptable standard for transmitting encrypted data.
- Encrypted E-mail – If secure FTP capabilities do not exist, encrypted e-mail can be used.
- Password Protected E-mail – If compatible encryption is not available to both parties, data should be password protected. The password should be given to the recipient through a different medium, such as a phone call, never in notes or documents accompanying the actual data file, or another e-mail. In addition, the password should not be transferred via voicemail.

When sending e-mail, either encrypted or password protected, it is advised to ensure that it contains the least amount of Family Educational Rights and Privacy Act (FERPA) –protected information as possible. The subject line of an e-mail should not include FERPA–protected information; the body of an e-mail should not contain highly sensitive FERPA–protected information, such as a student’s Social Security Number or full name. FERPA– protected data should always be in an attached encrypted/password protected file, never in the body of an email. Secure test questions, answer choices or portions of secure test questions or answer choices must not be sent via e-mail (use e-mail only if encrypted and/or password protected).

Fax machines and printers used to send and receive secure data must be located in areas that are secure. LEAs and charter schools should not use private or personal accounts to store students’ PII. LEAs and charter schools who wish to use the G suite for Education (previously called Google Apps for Education) should consult with their legal team to ensure compliance with FERPA and state security guidelines. Furthermore, it is recommended that the Data Leak Protection (DLP) feature of G Suite be used to protect data, even though FERPA compliance does not require DLP.

8.3.2 Confidentiality of Test Data

Confidential data must be transferred using secure methods (e.g., Secure File Transfer Protocol or receipted parcel delivery services, such as the U.S. Postal Service, UPS, or Federal Express).

When placing confidential data on portable devices (e.g., laptops, thumb drives), the portable device must be protected by encryption or password protection. Some specific examples of confidential data that must not be released to anyone include the following:

- WinScan files contain data that are for test development and accountability purposes only, and their release would violate test security.
- The EDS data are property of the NCDPI and School Nutrition Services. Accountability Services has access to the data through a Memorandum of Understanding (MOU). Test coordinators are bound by the requirements of the MOU and FERPA to preserve the confidentiality of this data. Releasing this data to anyone in any manner that would allow the identification of the EDS status of an individual student would be a violation of federal law.

CHAPTER 9 VALIDITY EVIDENCES

This chapter presents additional validity evidence collected in support of the interpretation of *Edition 4* mathematics grades 3–8 and NC Math 1 NCEXTEND1 Alternate Assessment test scores. The first two sections present validity evidence in support of the internal structure of the NCEXTEND1 assessments. Evidence presented in these sections include reliability, standard error estimates, and classification consistency summary of reported achievement levels and an exploratory principal component analysis (PCA) to support the unidimensional interpretation of the NCEXTEND1 mathematics scores. The final section of the chapter documents procedures used to ensure *Edition 4* NCEXTEND1 Alternate Assessments are accessible and fair for all students.

9.1 Reliabilities of the NCEXTEND1 Assessments

Internal consistency, as a reliability estimate, provides a sample base summary statistic that describes the proportion of the reported score variability that is attributed to true score variance. To justify valid use of test results in large-scale standardized assessments, evidence must be documented that shows test results are stable, consistent, and dependable across all subgroups of the intended population. A reliable assessment produces scores that are expected to be relatively stable if the test is administered repeatedly under similar conditions to the same students. Scores from a reliable test reflect examinees estimated expected ability in the construct being measured with very little error variance. Cronbach alpha as a measure of internal consistency ranges from 0.0 to 1.0, where a coefficient of 1.0 refers to a perfectly reliable measure with no measurement error. For high-stakes assessments, alpha estimates of 0.85 or higher are generally desirable. Cronbach’s alpha (Cronbach, 1951) is calculated as:

$$\hat{\alpha} = \frac{\kappa}{\kappa-1} \left(1 - \frac{\sum \hat{\sigma}_i^2}{\hat{\sigma}_X^2} \right) \quad (9-1)$$

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. It is worth noting that reliability estimates are less informative in describing the accuracy of individual students’ scores, since they are sample based. *Table 9.1* shows Cronbach alpha as a reliability coefficient for all mathematics NCEXTEND1 alternate assessments by grade and major demographic variables for 2018–19 administration. Across grades, reliability estimates based on the 2018–19 population ranged from the lowest of 0.72 in grade 7 to the highest of 0.87 in grade 5. Subgroup reliabilities are also in the similar range with the lowest for ELs population. Note that the total alternate student population was around 1,000 with smaller populations for subgroups. Furthermore, the subgroup population was not diverse in terms of student ability resulting in lower reliabilities.

Table 9.1 Overall and Subgroup Reliabilities, NCEXTEND1 Mathematics

Grade	Gender		Ethnicity ¹			Accommodations ¹			All SEM	
	Female	Male	Black	Hispanic	White	EDS	SWD	ELs		
3	0.80	0.84	0.84	0.78	0.84	0.82	0.83	0.77	0.83	3.9
4	0.80	0.81	0.80	0.80	0.81	0.81	0.80	0.78	0.81	3.9
5	0.87	0.86	0.88	0.81	0.88	0.85	0.87	0.79	0.87	3.1
6	0.80	0.82	0.78	0.77	0.83	0.81	0.81	0.74	0.81	3.9
7	0.72	0.72	0.74	0.63	0.72	0.72	0.73	0.70	0.72	4.0
8	0.77	0.81	0.78	0.78	0.82	0.80	0.80	0.82	0.80	4.0
10	0.76	0.75	0.67	0.78	0.79	0.73	0.75	0.78	0.75	4.1

¹Reliabilities estimates are displayed only for major ethnic groups and accommodations investigated in DIF analysis with acceptable sample size.

9.2 Conditional Standard Errors at Scale Score Cuts

The information provided by the standard error (SE) for a given cut score is important because it helps in determining the accuracy of examinees’ classifications. It allows a probabilistic statement to be made about an individual’s test score. The conditional SEs at the lowest obtainable scale score (LOSS), highest obtainable scale score (HOSS) and scale score cuts at the achievement levels for the mathematics NCEXTEND1 Alternate Assessments are shown in Table 9.2.

The conditional SE can be used to estimate a confidence band around any scale score or cut score where a decision must be precise. For example, the on-grade level proficiency (Level 3) cut score for grade 3 mathematics is 451 (see Table 9.2). A student scored 451 will have 68% probability that his or her true score or ability ranges from 447 to 455 ($451 \pm 1 \times 3.7$) when reported with one SE level of precision. Note that the SE for the Level 3 is 3.7. Similarly, if an educator wants to estimate students’ true score with less precision but higher confidence, say 95% confidence interval, it is given by two standard error above and beyond the cut or 444 to 458 ($451 \pm 2 \times 3.7$) depending on rounding rules.

The Level 3 scale score cuts for most of the mathematics grades 3–8 and NC Math 1 NCEXTEND1 Alternate Assessments ranged from 3 to 5, and Level 4 ranged from 2 to 3. The higher SE at the Not Proficient level is typical for extreme scores which allow less measurement precision because of a lack of enough informative items at those ability ranges.

Table 9.2 NCEXTEND1 Conditional Standard Errors at Achievement Level Cuts

Grade	Not Proficient		Level 3		Level 4	
	Cut	SE	Cut	SE	Cut	SE
3	419–450	6.7	451–463	3.7	464–470	1.7
4	419–450	6.6	451–464	3.7	465–474	2.4
5	417–451	5.6	452–464	3.9	465–475	2.7
6	419–452	7.2	453–463	2.9	464–474	2.7
7	414–449	5.8	450–466	4.6	467–477	2.7
8	418–452	6.4	453–464	3.5	465–473	2.4
NC Math 1	414–451	7.5	452–462	3.2	463–475	2.5

9.3 Classification Consistency

The *No Child Left Behind Act of 2001* (USDE, 2002) and subsequent *Race to the Top Act of 2009* (USDE, 2009) emphasized the measurement of AYP with respect to the percentage of students at or above performance standards set by states. With this emphasis on the achievement level classification, it is very important to provide evidence that shows all students are consistently and accurately classified into one of the four achievement levels. The importance of classification consistency as a measure of the categorical decisions when the test is used repeatedly has been recognized in Standard 2.16 (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. The 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 classification consistency analysis was conducted using the computer program BB-Class². The program provides results for both the Hanson and Brennan, or HB, (1990) and Livingston and Lewis, or LL, (1995) procedures. Since the Hanson and Brennan (1990) procedures assume “test consists of n equally weighted, dichotomously-scored items,” while the Livingston and Lewis (1995) procedures intends to handle situations when “a) items are not equally weighted and/or b) some or all of the items are polytomous scored” (Brennan, 2004, pp. 2–3), therefore the classification consistency analyses for the mathematics grades 3-8 and NC Math 1 NCEXTEND1 Alternate Assessments followed the HB procedures.

Error! Reference source not found. shows the decision accuracy and consistency indexes for the NCEXTEND1 Alternate Assessments’ achievement levels at each grade. Two observations can be made from the Table: 1) the accuracy and consistency values are consistent in the two levels, and 2) the accuracy and consistency values are higher (0.92–0.96) for the Level 4 compared to Level 3 (0.76 to 0.89). One can interpret the results as, if grade 3 NCEXTEND1 mathematics students who were classified as Level 3 were to take a non-overlapping, equally difficult form a second time, 85% (**bolded**) of them would still be classified as Level 3 and 92% (**bolded**) of them would still be classified as Level 4. Smaller SE seen at *Table 9.2* translates to higher reliability measures that will exhibit higher levels of classification consistency.

Table 9.3 Classification Accuracy and Consistency Results, NCEXTEND1 Mathematics

Grade	Level 3		Level 4	
	Sufficient Understanding Grade Level Proficient		Thorough Understanding Competitive Employment	
	Accuracy	Consistency	Accuracy	Consistency
3	0.89	0.85	0.93	0.92
4	0.88	0.83	0.94	0.93
5	0.88	0.83	0.95	0.94
6	0.87	0.82	0.94	0.93
7	0.85	0.80	0.96	0.95
8	0.88	0.83	0.93	0.93
Math I	0.83	0.76	0.95	0.93

² 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>.

9.4 Unidimensionality of NCEXTEND1 Assessments

The North Carolina mathematics grades 3–8 and NC Math 1 NCEXTEND1 Alternate Assessments are designed base on a unidimensional assumption that total score represents an estimate of students’ performance based on grade level mathematics content standards. It is therefore important that the NCDPI test design show relevant validity evidence to support the unidimensional use and interpretation of test scores.

Empirical evidence of overall dimensionality for the NCEXTEND1 mathematics assessments was explored using principle component analysis (PCA). The PCA is an exploratory technique that seeks to summarize observed variables using fewer linear dimensions, referred to as components. The primary hypothesis in a PCA is to determine the fewest reasonable dimensions or components that can explain most of the observed variance in the data. Two commonly used criteria to decide the number of meaningful dimensions for a set of observed variables are:

- retain components whose eigenvalues are greater than the average of all the eigenvalues, which is usually 1 and
- plot eigenvalues (scree plot) against components (factors) and count the number of components above the natural linear break.

It is very common to rely on both criteria when evaluating the number of possible dimensions for a given variable. PCA were extracted from the polychoric correlation matrix for categorical scored responses, to determine the number of meaningful components.

9.4.1 Eigenvalues and Variance

The eigenvalue for each component describes the amount of total variance accounted for by that component. A scree plot is used to show the graphical result from PCA showing the relations between main components and cumulative variance explained. *Figure 9.1* through *Figure 9.7* show the PCA results for all mathematics grades 3–8 and NC Math 1 NCEXTEND Alternate Assessment forms. The left vertical axis shows the actual eigenvalues of parallel forms, and the right vertical axis displays the cumulative variance.

Evaluation of the scree plots with the distinct break of the linear trend after the first dominant component present enough exploratory evidence in support of the assumption of unidimensionality with a single dominant component to explain a significant amount of the total variance of the mathematics grades 3–8 and NC Math 1 NCEXTEND1 Alternate Assessments.

Figure 9.1 Grade 3 NCEXTEND1 Mathematics Scree Plots of 2018–19 Operational Forms

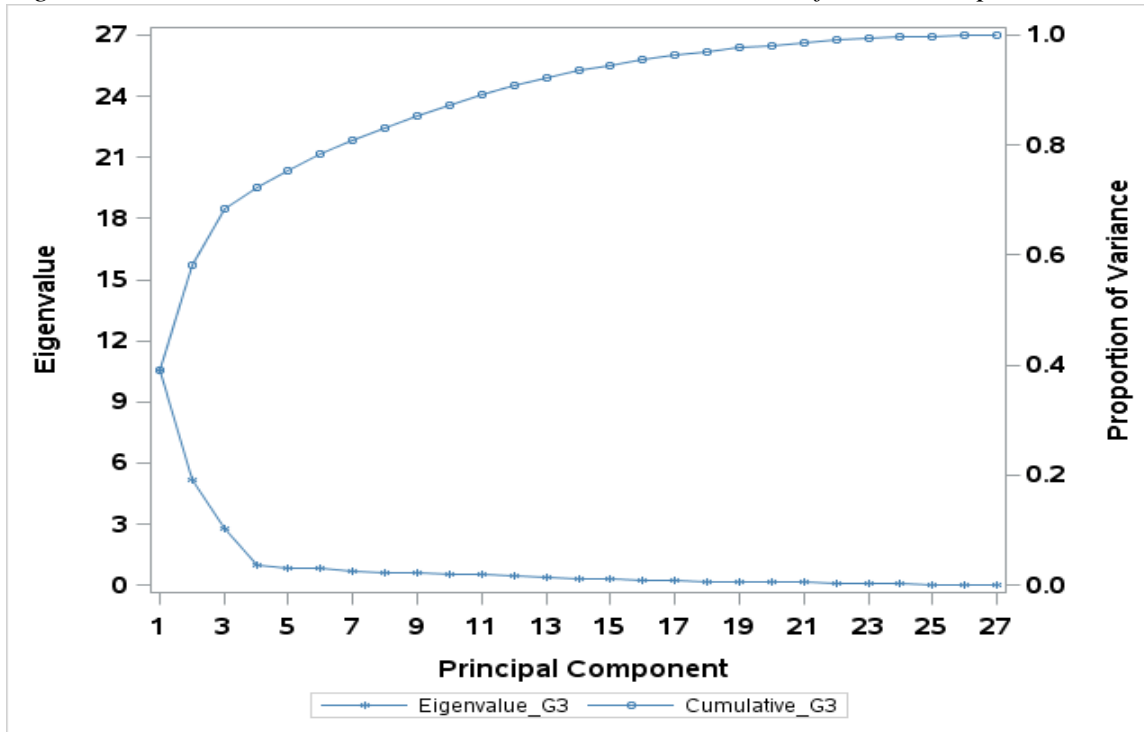


Figure 9.2 Grade 4 NCEXTEND1 Mathematics Scree Plots of 2018–19 Operational Forms

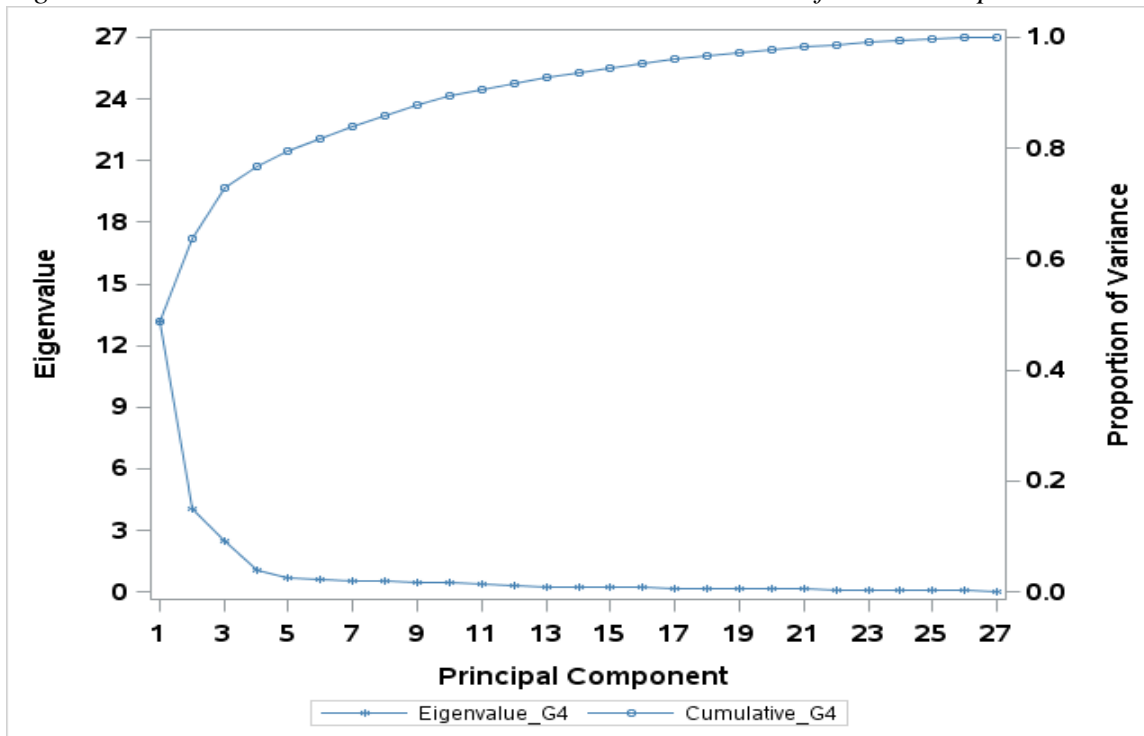


Figure 9.3 Grade 5 NCEXTEND1 Mathematics Scree Plots of 2018–19 Operational Forms

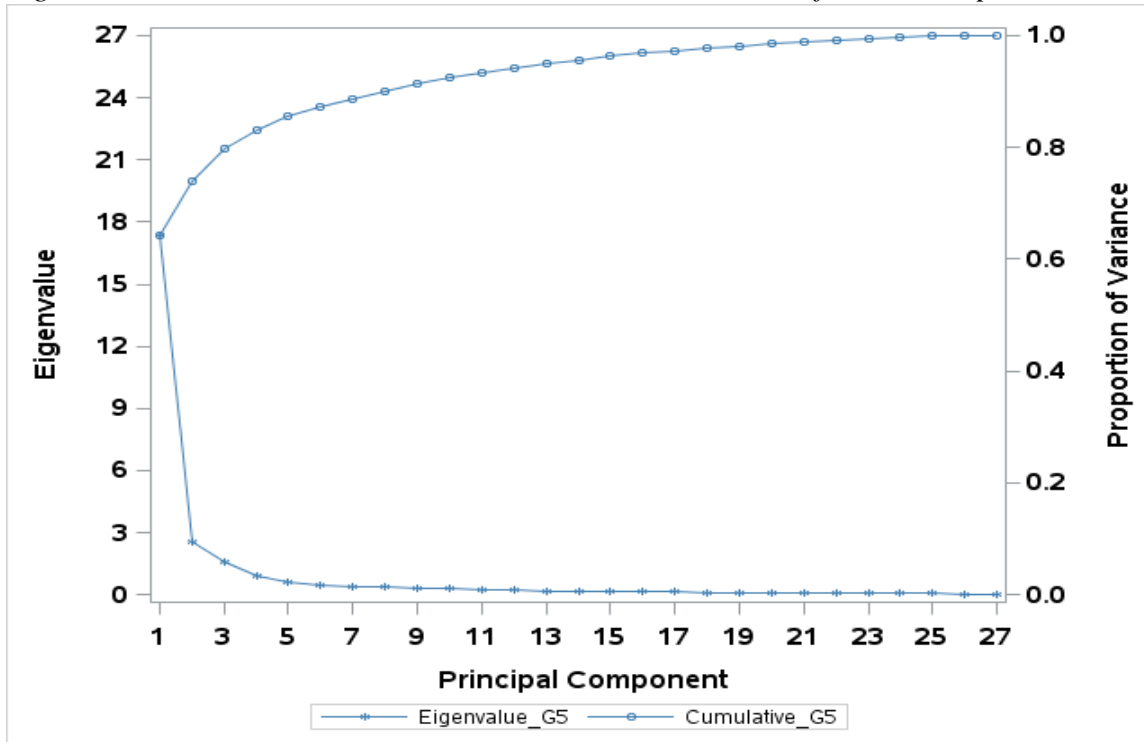


Figure 9.4 Grade 6 NCEXTEND1 Mathematics Scree Plots of 2018–19 Operational Forms

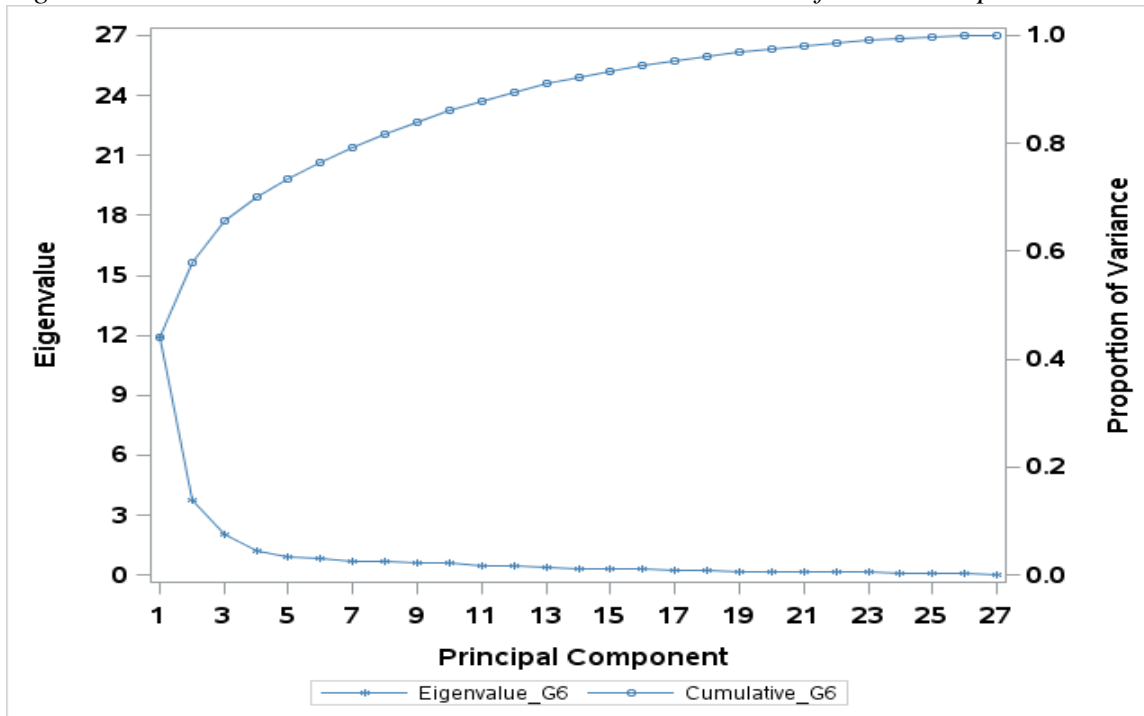


Figure 9.5 Grade 7 NCEXTEND1 Mathematics Scree Plots of 2018–19 Operational Forms

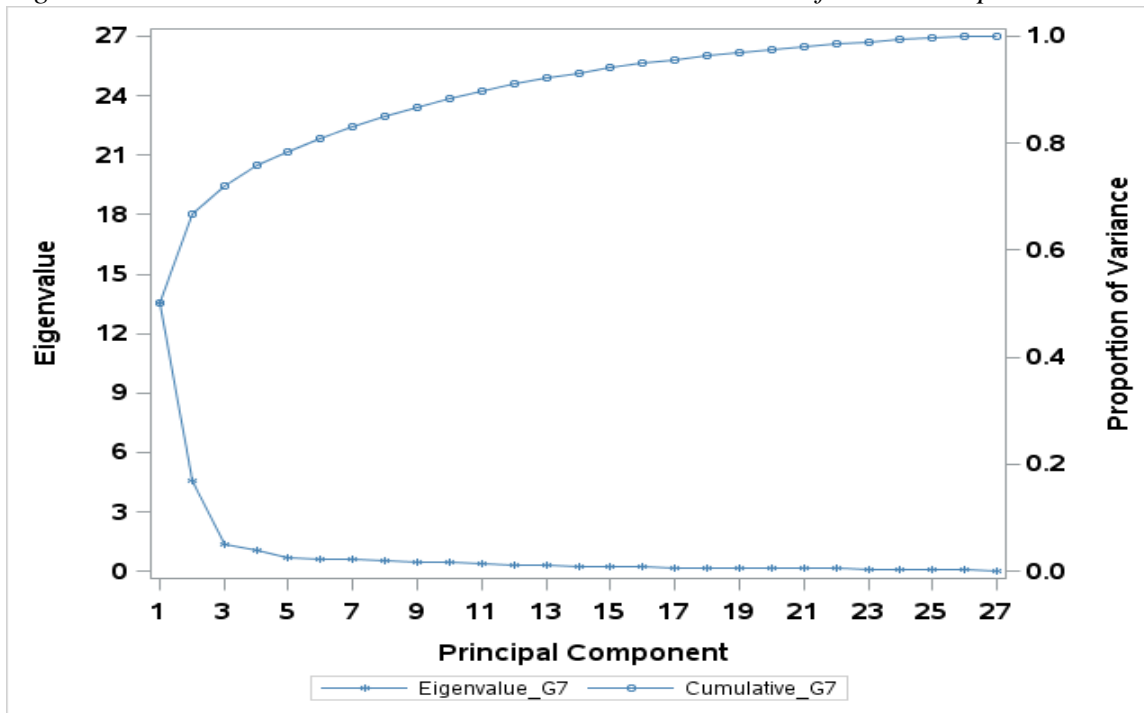


Figure 9.6 Grade 8 NCEXTEND1 Mathematics Scree Plots of 2018–19 Operational Forms

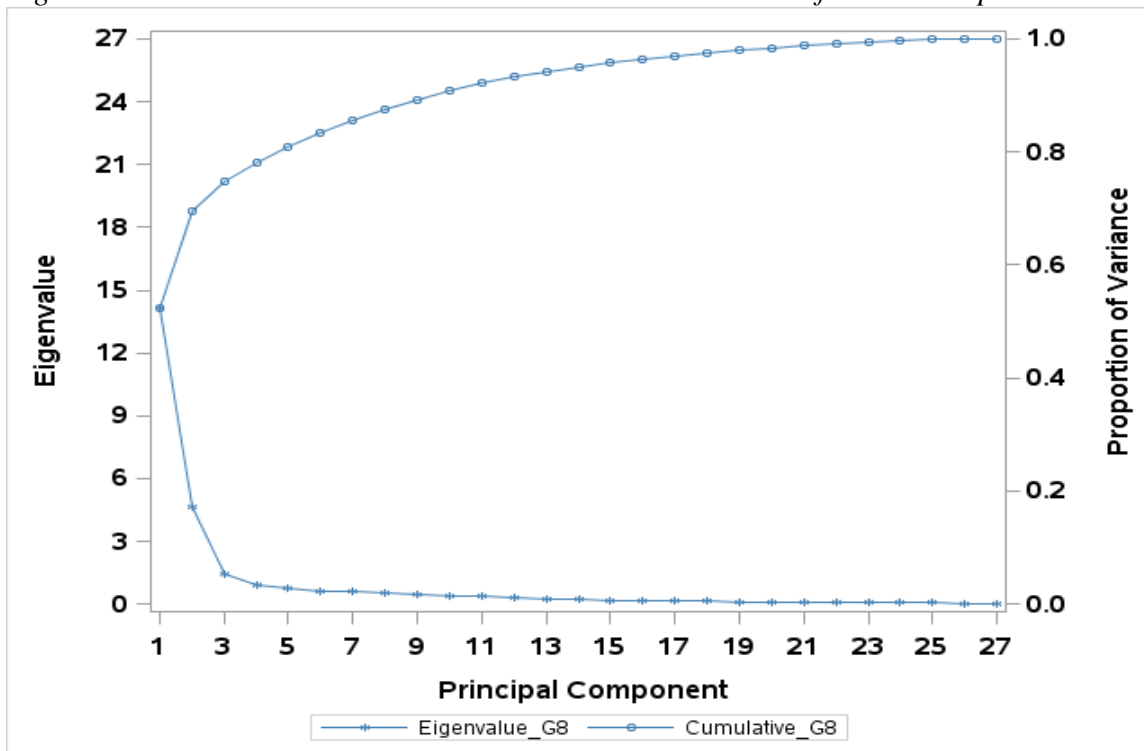
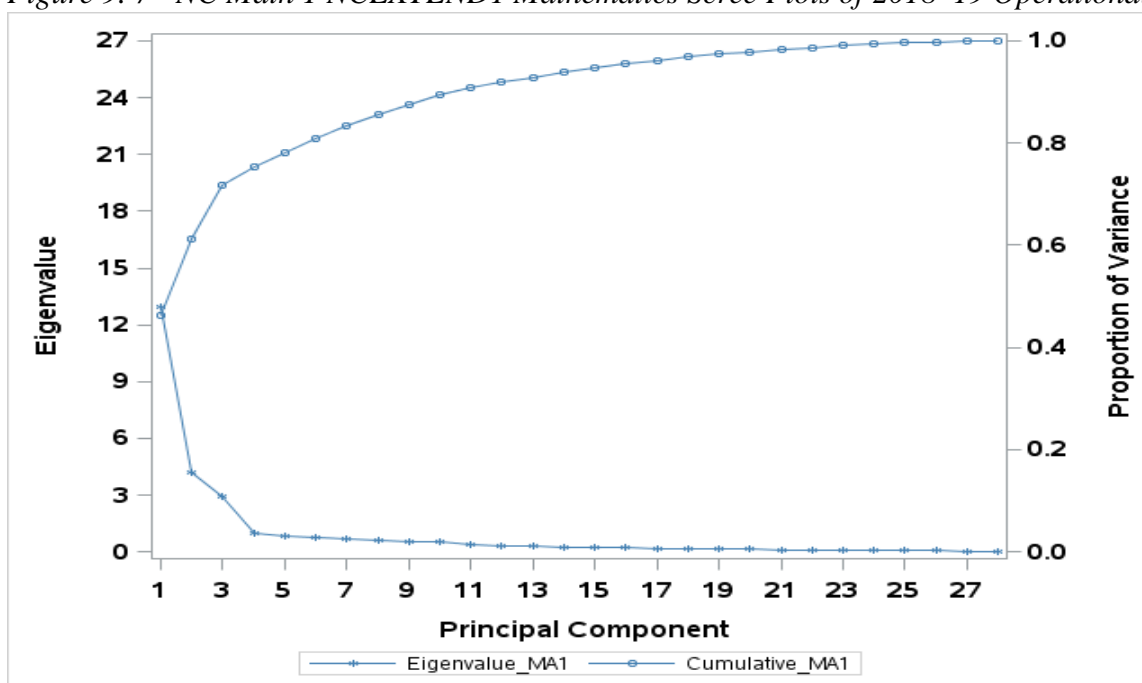


Figure 9.7 NC Math 1 NCEXTEND1 Mathematics Scree Plots of 2018–19 Operational Forms



The Eigen values and variance accounted for by the components for the first three components with Eigenvalues greater than 1 are summarized in *Table 9.4*. Based on the PCA results, the average ratio of the first to second eigenvalue across grades ranged from a minimum of approximately 2.0 in grade 3 to a maximum of 6.8 in grade 5 indicating significantly smaller proportion of variance explained by the second factor. Also, on average the first principal component accounted for about the lowest of 39% in grade 3 to the highest 64.4% in grade 5.

Table 9.4 Eigenvalues (Eigen) and Variance (%) Accounted for by the Components, Grades 3–8 and NC Math 1

Factor	Grade													
	3		4		5		6		7		8		Math 1	
	Eigen	%	Eigen	%	Eigen	%	Eigen	%	Eigen	%	Eigen	%	Eigen	%
1	10.5	39.0	13.1	48.7	17.4	64.4	11.9	44.0	13.5	50.0	14.2	52.4	12.9	46.2
2	5.2	19.2	4.0	14.9	2.6	9.5	3.8	14.0	4.5	16.8	4.6	17.2	4.2	15.1
3	2.8	10.4	2.5	9.2	1.6	5.9	2.0	7.6	1.4	5.1	1.4	5.3	3.0	10.6
4	1.0	3.8	1.1	3.9			1.2	4.4	1.0	3.8			1.0	3.5

9.5 Alignment Study

Alignment in large scale assessment refers to how well the assessment items and the assessment framework as a whole reflected the intended academic content and performance standards on which they are based. The collection of alignment evidence for the North Carolina assessments started from the item writing and test development phase where TMSs from NCSU-TOPS and the NCDPI as well as Psychometricians were responsible for training item writers for writing items aligned to academic content standards, selection of items representing test blueprint, performance expectations in terms of cognitive complexities or DOKs and creating a test reflecting target difficulty.

A formal alignment study quantifying the degree of alignments in the major outcome variables is planned for summer of 2022 administration.

9.6 Fairness and Accessibility

9.6.1 Accessibility in Universal Design

To ensure fairness and accessibility for all eligible students for the mathematics grades 3-8 and NC Math 1 NCEXTEND1 Alternate Assessments, the principle of universal design was embedded throughout the development and design of the tests. The NCEXTEND1 assessments measure student’s knowledge as defined in the *North Carolina Extended Content Standards*. Assessments 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, the NCDPI train item writer and reviewers with “Plain English Principles”.

Evidence of universal design principles applied in the development of the 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 this report. Some of the universal design principles used in the training include:

- Precisely defined constructs
 - Direct match to objective being measured
- Accessible, nonbiased items³
 - Accommodations included from the start (Braille, large–print, oral presentation etc.)
 - Ensuring that quality is retained in all items
- Simple, clear directions and procedures

³ See discussions on fairness review in Chapter 4

- Presenting in understandable language,
- Using simple, high frequency and compound words,
- Using words that are directly related to content the student is expected to know,
- Omitting 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
- 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 ELs Students
 - Test in a separate room
- Computer-based Forms
 - All students receive one item per test page,
 - All students may receive larger font and different background colors.

9.6.2 Fairness in Access

Alignment evidence, presented throughout Chapter 2 through Chapter 6, demonstrated the NCDPI’s commitment that all assessment blueprints are aligned to content domains that are also aligned to the NCSCS. Assessments’ content domain specifications and blueprints are published on the NCDPI public website with other relevant information regarding the development of Alternate 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 mathematics grades 3–8 and NC Math 1 NCEXTEND1 Alternate Assessments, the NCDPI also published released items for every grade level, which were constructed using the same blueprint as the operational forms.

These released items provided students, teachers and parents with sample items and a general practice form that is similar to the operational assessment. These released items also served as a resource to familiarize students with the various response formats in the new assessments.

9.6.3 Fairness in Administration

Chapter 5 of this report documents the procedures put in place by the NCDPI to assure that the administration of the NCEXTEND1 assessments are standardized, fair and secured for all students across the state. For each assessment, the NCDPI publishes a *North Carolina Test Coordinators' Policies and Procedures Handbook* and *Assessor's Guide* that are the main training materials for all assessors across the state. These documents provide comprehensive details of policies and procedures for each assessment including general overview of each assessment that covers the purpose of the assessment, student eligibility, testing window, and makeup testing options. Assessor's Guides also covers all preparations and steps that should be followed the day before testing, on test day, and after testing. Samples of answer sheets are also provided in the guide.

The NCDPI recommends that the online assessment tutorial should be used to determine students' appropriate font size (i.e., regular or large) and/or alternate background color for test day. These options must be entered in the student's interface questions before test day. The online assessment tutorial can assist students, whose IEP or Section 504 Plan designates the Large Print accommodation in determining, whether the large font will be adequate for the student on test day. If the size of the large font is insufficient for a student because of his/her disability, this accommodation may be used in conjunction with the *Magnification Devices* accommodation, or a *Large Print Edition* of the paper-and-pencil assessment may be ordered.

In order to prepare students in their NCEXTEND1 mathematics tests, the NCDPI produced practice activities and required students take the practice activity before the administration of the tests. Schools must ensure that every student participated in the practice activity at least one time at the school before the test day.

Glossary of Key Terms

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

Key Terms	Definition
Accommodations	Changes made in the format or administration of the test to provide options to test takers who are unable to take the original test under standard test conditions.
Achievement Levels	Descriptions of a test taker’s competency in a particular area of knowledge or skill, usually defined as ordered categories on a continuum classified by broad ranges of performance.
Biserial Correlation	The relationship between an item score (right or wrong) and a total test score.
Cut Scores	A specific point on a score scale, such that scores at or above that point are interpreted or acted upon differently from scores below that point.
Dimensionality	The extent to which a test item measures more than one ability.
Embedded Field-Test Design	Using an operational test to FT new items or sections. The new items or sections are “embedded” into the new test and appear to examinees as being indistinguishable from the operational test.
Equivalent Forms	The differences between forms are not statistically significant.
Field-Test	A collection of items to approximate how a test form will work. Statistics produced will be used in interpreting item behavior/performance and allow for the calibration of item parameters used in equating tests.
Foil Counts	Number of examinees that endorse each foil (e.g., number who answer “A,” number who answer “B,” etc.).
Operational Test	Test administered statewide with uniform procedures, full reporting of scores and stakes for examinees and schools.
P-value	Difficulty of an item defined by using the proportion of examinees who answered an item correctly.

Key Terms	Definition
Parallel Forms	Forms that are developed with the same content and statistical specifications.
Percentile	The score on a test below which a given percentage of scores fall.
Raw Score	The unadjusted score on a test determined by counting the number of correct answers.
Scale Score	A score to which raw scores are converted by numerical transformation. Scale scores allow for comparison of different forms of the test using the same scale.
Standard Error of Measurement	The standard deviation of individuals' observed scores, usually estimated from group data.
Test Blueprint	The testing plan, which includes the numbers of items from each objective that are to appear on a test and the arrangement of objectives.

References

- AERA, APA, & NCME (2014). Standards for educational and psychological testing. Washington, D.C.: Author.
- Angoff, W.H. (1971). Scales, norms and equivalent scores. In R.L. Thorndike, (Ed.), Educational Measurement, 2 nd Ed. Washington, DC: American Council on Education, 508-600.
- 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).
- 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.
- 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.
- Hambleton, R. K. (2001). Setting performance standards on educational assessments and criteria for evaluating the process. In G. J. Cizek (Ed.), *Setting performance standards: Concepts, methods, and perspectives* (pp. 89–116). Mahwah, NJ: Lawrence Erlbaum.
- Hanson, B.A. & Hess, K. (2013). *Cognitive Rigor in Today’s Classroom-Using a Cognitive Rigor Matrix to Advance Complexity of Thoughts*. Measured Progress, 2013.
- 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.
- Lewis, D. M., Green, D. R., Mitzel, H.C., Baum, K. & Patz, R.J. (1998). The Bookmark standard setting procedure: Methodology and recent implementations. Paper presented at the annual meeting of the National Council on Measurement in Education. San Diego, CA.
- 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. S., & Cizek, G. J. (2012). Variations on a theme: The modified Angoff, extended Angoff, and yes/no standard setting methods. In G. J. Cizek (Ed.) *Setting performance standards: Foundations, methods, and innovations*(pp. 181-200) New York: Routledge.
- SAS Institute, Inc. (1985). *SAS User's Guide: Statistics, Version 5 Edition*. Cary, NC: Author.

USED (2009) Race to the Top Program Executive Summary. U.S. Department of Education, Washington, D.C. 20202

Zieky, M. J. (2012). So much has changed. In G. J. Cizek (Ed.), *Setting performance standards: Foundations, Methods, and Innovations*, 2 nd Ed. New York: Routledge, 15-32.

Appendix 1-A

Testing Code of Ethics

<https://www.dpi.nc.gov/media/119/open>

Appendix 2–A

Math Test Specification Meeting Agendas, Survey Form, and Demographic Information of Participants

Test Specification Meeting

DAY 1—Meeting Agenda

North Carolina Department of Public Instruction/Room 150 North

8:30am	Registration—Room 150 North Betty Barbour, Josh Griffin
9:00am	Welcome and Introductions Josh Griffin, Hope Lung, Betty Barbour <ul style="list-style-type: none"> • Internet Access, Restrooms and Café (cash only) • Substitute Teacher Form, Stipend Form, Demographics Form • Testing Code of Ethics and Test Security Agreement • Travel Reimbursement
9:35am	Summative Assessment Psychometric Overview Dr. Kinge Mbella, Lead Psychometrician, NCDPI/Test Development
10:20am	Break
10:30am	Overview of Revised Standards for Math Grades 3-5 Kitty Rutherford and Denise Shultz, NCDPI/K-12 Mathematics Curriculum and Instruction
11:45am	Lunch (on your own)
12:45pm	Prioritize Standards—ROUND 1 (Breakout Groups: Grade 3, Grade 4, Grade 5) Josh Griffin, Math Test Measurement Specialist, NCDPI/Test Development <ul style="list-style-type: none"> • Prioritize Assessable Standards • Recommend Weighting by Domain
	Break (on your own)
2:00pm	Prioritize Standards—ROUND 2 (Large Group) Josh Griffin <ul style="list-style-type: none"> • Prioritize Assessable Standards • Recommend Weighting by Domain
3:00pm	Recommend Percent by Item Type, Calculator Use—Discussion (Large Group) Josh Griffin
3:45pm	Summary of Recommendations and General Considerations Josh Griffin
4:00 pm	Meeting Adjourned (Bring snacks/drinks for Day 2) Josh Griffin

DAY 2—Meeting Agenda

9:00am	Collect Travel Reimbursement Documentation Betty Barbour
9:15am	Overview of Cognitive Complexity (Webb’s Depth of Knowledge) Josh Griffin, Math Test Measurement Specialist, NCDPI/Test Development

	Break (on your own)
11:15am	Cognitive Complexity—ROUND 1 (Breakout Groups: Grade 3, Grade 4, Grade 5) Josh Griffin <ul style="list-style-type: none"> • Recommend Percent by DOK Level
	Break (on your own)
12:30pm	Cognitive Complexity—ROUND 2 (Large Group) Josh Griffin <ul style="list-style-type: none"> • Recommend Percent by DOK Level
1:00 pm	Distribution of Certificates and Meeting Adjourned Josh Griffin

Demographic Form Test Specifications Meeting

Purpose: The completion of this form is voluntary. We are requesting information from each individual because it will provide a description of this group. This information will be used by the North Carolina Department of Public Instruction for aggregate data analysis only. Thank you for your consideration!

Information

(Optional) Print your Name: _____

Gender: Male Female

Ethnicity: _____

Education

Highest Degree Earned: B.A./B.S M.A./M.S./M.Ed. Ed.D/Ph.D Other:

Approximate Year Highest Degree Received: _____

Experience

(Active teachers only) What grade level(s) or course(s) did you teach in 2016–17?

National Board Certified (circle one): Yes No

If Yes, list your National Board Certification Fields:

North Carolina Teacher Certification Fields:

Number of Years Employed in Education: _____

Grade Levels Taught (include your entire teaching career; circle all that apply):

K 1 2 3 4 5 6 7 8 9 10 11 12

Experience Teaching the Following (circle all that apply):

EL Students Students with Disabilities Gifted Students Extended Content
Standards

Employment

Employment Classification (circle one): Full-Time Part-Time Retired

If Full-Time or Part-Time, what is the title of your position? _____

Are you employed by a charter school (circle one)? Yes No

If YES, what is the name of the charter school?

Are you employed by a school district (circle one)? Yes No

If YES, what is the name of the school district?

If you work at the school-level, what is the name of the school?

Compared to other school districts in North Carolina, which of the following best describes the size of your district (meaning the number of students attending schools in your district)?

Large

Medium

Small

Compared to other school districts in North Carolina, which of the following best describes the community setting of your district (circle one)?

Urban

Suburban

Rural

Table 2.1 Demographic Characteristics of the Test Specification Meeting Participants

Category	Sub-Category	NC Math 2 & 3 (N=24)		Grade 3-5 (N=40)		Grade 6-8 (N=39)		NC Math 1 (N=13)	
		N	%	N	%	N	%	N	%
Gender	Female	20	83%	36	90%	34	87%	11	85%
	Male	4	17%	4	10%	5	13%	2	15%
Ethnicity	Asian		0%	1	3%		0%		0%
	Black	1	4%	5	13%	7	18%	4	31%
	Native American		0%		0%	1	3%		0%
	Hispanic	1	4%		0%		0%		0%
	White	8	33%	32	80%	25	64%	9	69%
	Mixed		0%	1	3%	1	3%		0%
Highest Degrees Earned	BA/BS	7	29%	14	35%	15	38%	5	38%
	J.D./Ed.D/Ph.D	2	8%	4	10%	1	3%	2	15%
	MA/MS/M.Ed	15	63%	22	55%	23	59%	6	46%
District Size	Large	10	42%	13	33%	16	41%	6	46%
	Large/Medium							1	8%
	Medium	5	21%	14	35%	13	33%	2	15%
	Small	5	21%	8	20%	5	13%	2	15%
Urbanicity	Rural	8	33%	14	35%	8	21%	3	23%
	Suburban	5	21%	11	28%	12	31%	5	38%
	Suburban/Rural	1	4%	2	5%	2	5%		0%
	Urban	4	17%	6	15%	5	13%	2	15%
	Urban/Suburban		0%	1	3%	2	5%	1	8%
	Urban/Suburban/Rural		0%	1	3%	1	3%		0%

*Some participants did not declare some of the demographic characteristics

Appendix 2–B

Mathematics Depth-of-Knowledge Levels

Mathematics Depth-of-Knowledge Levels

Level 1 (Recall) includes the recall of information such as a fact, definition, term, or a simple procedure, as well as performing a simple algorithm or applying a formula. That is, in mathematics a one-step, well-defined, and straight algorithmic procedure should be included at this lowest level. Other key words that signify a Level 1 include “identify,” “recall,” “recognize,” “use,” and “measure.” Verbs such as “describe” and “explain” could be classified at different levels depending on what is to be described and explained.

Level 2 (Skill/Concept)

Level 2 includes the engagement of some mental processing beyond a habitual response. A Level 2 assessment item requires students to make some decisions as to how to approach the problem or activity, whereas Level 1 requires students to demonstrate a rote response, perform a well-known algorithm, follow a set procedure (like a recipe), or perform a clearly defined series of steps. Keywords that generally distinguish a Level 2 item include “classify,” “organize,” “estimate,” “make observations,” “collect and display data,” and “compare data.” These actions imply more than one step. For example, to compare data requires first identifying characteristics of the objects or phenomenon and then grouping or ordering the objects. Some action verbs, such as “explain,” “describe,” or “interpret” could be classified at different levels depending on the object of the action. For example, if an item required students to explain how light affects mass by indicating there is a relationship between light and heat, this is considered a Level 2. Interpreting information from a simple graph, requiring reading information from the graph, also is a Level 2. Interpreting information from a complex graph that requires some decisions on what features of the graph need to be considered and how information from the graph can be aggregated is a Level 3. Caution is warranted in interpreting Level 2 as only skills because some reviewers will interpret skills very narrowly, as primarily numerical skills, and such interpretation excludes from this level other skills such as visualization skills and probability skills, which may be more complex simply because they are less common. Other Level 2 activities include explaining the purpose and use of experimental procedures; carrying out experimental procedures; making observations and collecting data; classifying, organizing, and comparing data; and organizing and displaying data in tables, graphs, and charts.

Level 3 (Strategic Thinking)

Level 3 requires reasoning, planning, using evidence, and a higher level of thinking than the previous two levels. In most instances, requiring students to explain their thinking is a Level 3. Activities that require students to make conjectures are also at this level. The cognitive demands at Level 3 are complex and abstract. The complexity does not result from the fact that there are multiple answers, a possibility for both Levels 1 and 2, but because the task requires more demanding reasoning. An activity, however, that has more than one possible answer and requires students to justify the response they give would most likely be a Level 3. Other Level 3 activities include drawing conclusions from observations; citing evidence and developing a logical argument for concepts; explaining phenomena in terms of concepts; and using concepts to solve problems.

Level 4 (Extended Thinking)

Level 4 requires complex reasoning, planning, developing, and thinking most likely over an extended period of time. The extended time period is not a distinguishing factor if the required work is only repetitive and does not require applying significant conceptual understanding and higher-order thinking. For example, if a student has to take the water temperature from a river each day for a month and then construct a graph, this would be classified as a Level 2. However, if the student is to conduct a river study that requires taking into consideration a number of variables, this would be a Level 4. At Level 4, the cognitive demands of the task should be high and the work should be very complex. Students should be required to make several connections—relate ideas *within* the content area or *among* content areas—and have to select one approach among many alternatives on how the situation should be solved, in order to be at this highest level. Level 4 activities include designing and conducting experiments; making connections between a finding and related concepts and phenomena; combining and synthesizing ideas into new concepts; and critiquing experimental designs.

Appendix 2–C

A Guide for Using Webb’s Depth of Knowledge with Common Core State Standards

[A Guide for Using Webb’s Depth of Knowledge with Common Core State Standards \(ohio.gov\)](#)

Appendix 2–D

NCEXTEND1 Alternate Assessment item development and review process

<https://www.dpi.nc.gov/media/9618/open>

Appendix 2–E

Alternate Assessment System Tryout Observation

Fall 2017 Math 1 Alternate Assessment System Tryout Observation Notes

Procedure:

One person from DPI and one from TOPS visited five high schools. Four schools were in Wake County and one was in Johnston County. The teachers were recruited by their district. Each teacher selected a student who was eligible to take the NC Math *EXTENDI* assessment.

Of the five high school teachers of students with disabilities observed, four were female and one was male. All the teachers appeared to be Caucasian, and had previously given the *NCEXTENDI* at least three times.

Of the students observed, two were female and three were male.

The following information was sent to the teachers prior to the observation visit.

Purpose:

The purpose of the system tryout is to assess how the online system works. This allows NCDPI to understand firsthand the functionality of the online system and gain your valuable feedback to help us make improvements.

The items on this test are not meant to measure a student's performance, but instead the items simply give us material to put on the online system so that we can observe how it works.

Process:

There are three administrations options for the system tryout:

1. Teacher facilitated online with paper manipulatives. The computer is positioned in front of the teacher and the directions are read from the computer screen to the student. The student sits across from the teacher (not viewing the screen). The cards are presented to the student while reading the script from the computer screen. The student indicates a single response using his/her reliable mode of communication. The student's responses are recorded on the computer by the teacher.
2. Teacher facilitated online. Both teacher and student sit facing the computer. The directions are read from the computer screen to the student. The student indicates a single response using his/her reliable mode of communication. The teacher records the student's responses on the computer. (Optional: The paper manipulative cards that coincide with the online test questions can also be placed in front of the student.)
3. Teacher facilitated online. Both teacher and student sit facing the computer. The directions are read from the computer screen to the student. The student enters his/her responses on the computer, using assistive technology if needed. (Optional: The paper manipulative cards that coincide with the online test questions can be placed in front of the student.)

After administering the tryout to the student, NCDPI staff will conduct a short debrief interview to gather your valuable feedback.

What is needed:

- Functioning computer/tablet with [insert whatever tech ability needed here]
- Computer/Device must have internet connection
- Space for working within one of the options above
- Any relevant assistive technology needed for the student to access a computer

Materials provided:

- Cards (provided by DPI) If you need to modify, please let us know, so that we can send the cards to you ahead of time.
- Log in instructions

Directions:

Directions for logging into the system will be provided on the day of the system try out.

What happens next:

Once the system tryouts are completed, we will review teacher feedback and proceed with any online system enhancements.

At the beginning of the observation visit, DPI and TOPS staff talked with the teacher about the background, purpose, and goals of the observation visit. A description of the system was provided, and a walk through prior to use with the student. After the first visit, ideas that were suggested in previous visits were talked about with each teacher.

The teacher was then asked to pick which method of administration (from choices listed above) would be used. Then system trial was conducted with the student. After the trial, the teacher was asked about the process and suggestions to improve the process and system.

Once all the visits were completed, the list of system recommendations were combined and placed into categories of need immediately and can wait. Suggestions for training and documentation were also included based on the observations.

System Recommendations to be implemented immediately

- Progress Indicator
 - Could be numbering the questions (1 of 10)
 - Could be a progress bar
- Clear division between part one and part two, directions for the administrator on how to continue or to comeback another day/time to finish.
 - Not everything has to be on the screen, but enough for them to understand what to do
 - Chance to pause or continue and directions on how to do the chosen response
 - This should be implemented during operational testing, not needed during the field test
- Tutorial/demo to walk through the system and to prepare the students

- One practice for all levels with three questions
- Allow teachers to practice as well as students
- Resume -1 (resume to the previous question)
 - This would void the answer coded incorrectly and reset the student to the previous question
 - The updated answer would then be entered in the system and scored
 - This would need to be indicated immediately (cannot go past the) and fixed on the spot, therefore was suggested that a STC be able to perform this function
 - Cleveland DTC agreed that the STC could perform this function during testing
- Clear directions provided (on screen) to make sure Assessor knows what to read
 - Might be on the login screen – read the bold/italic...
- Cards should be numbered so they can be put back in order
 - May not use all of them (just ones for graphs or things that need to be enlarged)
- Set **multiple test sessions** automatically for all students
- If going with stationary presentation and the item “appears” but does not disappear, need a way to enlarge answer choices, should also automatically provide cards for these choices
 - Magnifier that is large enough to click on without clicking on the answer

Recommendations to be added as possible

- Have two types of logins for teachers
 1. Teacher will use the cards to administer the test to the student – the screens are shortened with directions for the teacher to read and pick the answer for the student (don’t have to go through each screen – limit the number of clicks needed)
 2. Teacher will use the computer with the student – set up for viewing the question as it needs to be presented. (will have more clicks)

Training and Suggestions for administration:

- Need rules for projecting content onto smartboards or other devices.
 - Teachers need to know about what can be used –and how to maintain security
 - Teachers need help understand they can use all of the tech that they use in instruction
 - Who can help with administration – who can access the computer
- Cards would require two people for administration in current configuration
 - Need directions on who can access the computer/cards and how to do so
- Structured “casual” approach to directions
 - How to give structured directions that are not so formal
 - Options for how to tell a student to think about the question again and try it again.
- Make the questions more functional

- One- or two-pager for directions while using the computer
- Training needs to cover what to say, how to say it, where there is flexibility
 - What can you say to have the student try the question again in part one of the test
- Want some direction on how to pause/stop/restart the student
- Wanted to give some, but not all students, directions at the beginning – if you get it wrong, you will have a second chance at the items
 - May need to give directions at the start of part two that there is not a second chance at the item
- Will need to determine policy for students not completing the test.
 - Might need a comment section
- If reset -1 is implemented, need training to make the change/call immediately, do not proceed with the test until that question is reset
- Need to identify what the pause button does and when to use it – during breaks, to exit from the test ...
- Need to make sure teachers know that the test might end at the end of part one
 - The test is “adaptive” in that if a student is able to answer enough questions correctly they will go on, but if not, the test will end

Observations:

- **Athens Drive – November 14, 2017**
- **Enloe High School– November 20, 2017**
- **Panther Creek – November 21, 2017**
- **Fuquay Varina High School – November 29, 2017**
- **Cleveland High School – December 4, 2017**

Athens Drive Observation Notes November 14, 2017

Teacher 1 described self as having above average technology skills. The technology for the tryout was good – Windows 7 desktop. The student was described as using technology in the classroom – iPad and computer for both education and entertainment. Teacher 1 indicated that the student would want to click responses, but once we demonstrated the system, Teacher 1 wanted to be the one to click responses.

Teacher 1 chose to sit next to the student at the computer and for Teacher 1 to control the mouse. This would be the process used (out of the options provided) for all of Teacher 1’s current students. When asked to think about all previous students, Teacher 1 indicated that it would be the method used for about 90% of the students.

It was very easy for Teacher 1 to get started. Teacher 1 requested that the text for the administrator to read (directions) to be larger font. Cards were not used with the computer. Teacher 1 did not reread the question, so directions were needed to identify that question stems and answer choices are re-read when scaffolded.

Teacher 1 indicated that being on a computer was quicker than using the cards. Teacher 1 would need to control the mouse for most students. When given the option presented at the TAC of the items staying on the screen when introduced – Teacher 1 indicated that she would like that option. The no response button was good and in the right place, it reflects the student’s instruction. Teacher 1 would like to have a 1-pager for directions, but not a booklet to read from to go with the computer administration.

When asked about past administrations of the *NCEXTEND1*, Teacher 1 indicated that the cards were only adapted for visually impaired students. Teacher 1 has made them high contrast by darkening the lines and enlarged them. Teacher 1 was NOT in favor of colors, as differences in colors may distract the students.

Suggestions from Teacher 1:

- Progress bar – some way to know the beginning and end of the task.
- Teacher 1 would like a more “casual” approach to the directions. The stiff formal language distracted the students. (You will now be taking the...)
- Make the questions more functional – when you go to the store, which is the more expensive item?
- After the first set of questions (everyone gets), if the student will go on, need to be able to choose to continue (ask student) or to come back another day.
- Need to have the ability to use cards for the graphs, or complicated items.
- Need to identify what the pause button does

Suggestions from DPI after watching this administration:

- Clear division between part 1 and part 2, directions for the administrator on how to continue or to comeback another day/time to finish.
 1. Not everything has to be on the screen, but enough for Administrator to understand what to do.
- Have two types of logins for teachers
 1. Teacher will use the cards to administer the test to the student – the screens are shortened with directions for the teacher to read and pick the answer for the student (don’t have to go through each screen – limit the number of clicks needed).
 2. Teacher will use the computer with the student – set up for viewing the question as it needs to be presented. (will have more clicks)

Enloe Observation Notes – November 20, 2017

Teacher 2 indicated having a very good level of technology usage and knowledge. Teacher 2’s students use the computer every week to test. They have a laptop for every student. Did not use a mouse during the tryout.

During the observation, the wireless had some difficulty with one item – one response was delayed in appearing on the screen. It was delayed during the initial viewing of the answer choice, and also when the full question was on screen. It did appear after a few seconds.

Teacher 2 chose to control the item selection. Since the student was eager and willing, the exercise was conducted a second time with the student controlling the clicking. This would work if the system was identifying the correct response.

Teacher 2 indicated that 40% of current students would be able to do the process she demonstrated with this student. The other 60% might need some of the cards available (graphs, more complicated answer choices).

When the question and all information was on the screen, Teacher 2 did not repeat the stem for the item, just the new information on the screen (Choose an answer). When asking the student to look at an item again Teacher 2 said, “think about it again.” Teacher 2 might also say “Look at it one more time.” Teacher 2 questioned if the graph could be read for the student. This indicates that there needs to be clarity on the screen for what the Assessor will be read. Teacher 2 did not think that a progress bar was very important for students to have.

Suggestions from Teacher 2:

- Need to have the ability to use cards for the graphs, or complicated items.
- Some students might need the cards, but might not need to use all of the cards.
- Did not mind the way the item was presented – with each answer choice being separate.
 - This made the choices easier to see
 - The graphs would have been very difficult to see if only added to the page at the bottom
- Did think it would be a good idea to have 2 ways to log into the system – to eliminate some clicking for teachers only using cards with the student.
 - Teacher will use the cards to administer the test to the student – the screens are shortened with directions for the teacher to read and pick the answer for the student (don’t have to go through each screen – limit the number of clicks needed).
 - Teacher will use the computer with the student – set up for viewing the question as it needs to be presented. (will have more clicks)
- For compliance, wanted what Assessor was to say on the screen, could be on the bottom in *italics*
- Would be helpful to have a directions sheet when giving the test.
- Wants some direction on how to pause/stop/restart the student
- Wanted to give some, but not all students, directions at the beginning – if you get it wrong, you will have a second chance at the items
 - Will need to give directions at the start of part 2 that there is not a second chance at the item.
- Progress bar was not a priority

Suggestions, comments and notes from DPI and TOPS after watching this administration:

- Set multiple test sessions for all students automatically
 - Students may need to stop and start the assessment multiple times.
- Will need to determine policy for students not completing the test.
 - Might need a comment section
- Need to work on the pause between sections
- Need to identify what the pause button does
- Need to create a tutorial for this test
- If going with single page mode of presentation, need a way to enlarge some choices, or automatically provide cards for these responses.
 - Number the cards for ease of use and return

Panther Creek Observation – November 21, 2017

Teacher 3

Teacher 3 described self as exactly average level of technology knowledge and usage. Later when Teacher 3 did not understand “push F5,” joked that maybe rated self too high.

Teacher 2 indicated that the students use computers every day in class. They take an assessment monthly on the computer (same system as Enloe). The student we observed could use a cell phone, make some calls, and use it for some games.

Teacher 3 chose to use the computer and control the mouse. Teacher 3 indicated that the computer would be used with 100% of current students. The graph items would need cards. Most students would not need cards. Teacher 3 did not read the “teacher directions” at the top of the separation pages.

Teacher 3 used words like “try again” when the student needed to rethink the answer choice. Teacher 3 stated that this process seems more efficient and integrated. Teacher 3 indicated that one-pagers are effective with adults, so they would be more likely to be read than a manual.

Suggestions from Teacher 3:

- One-pager for directions for the Assessor
- Cards for the more complicated/complex items (e.g. graphs)
 - Need to be able to make graphs full screen
- Would like the fixed screen with reveal
- If don’t have instructions, will see inconsistent data – but would like to see several sets of directions
 - If you have an autistic student, say things this way
 - If you have a ... student, use these directions
- Did think it would be a good idea to have two ways to log into the system – to eliminate some clicking for teachers only using cards with the student.

Fuquay Varina HS Observation November 29, 2017

Teacher 4 indicated being good with technology, Teacher 4's father was an IT person. The computer was a Windows 7 machine using Chrome. The student has a 1:1 aide. The student routinely uses a smartboard and iPad in class and can use a cell phone for music. They use mostly paper and pencil with this student.

Teacher 4 decided to use option one and use the cards with the student. DPI helped place the cards in front of the student, while Teacher 4 used the computer. Teacher 4 indicated that in order to use the cards, there would have to be a second person helping. The rules around what the second person could do would need to be clear – who can touch the test materials. Teacher 4 indicated that 2 types of log in would be good. Teacher 4 indicated would have used the smartboard with this student, if we had been in the classroom. 10 of her 12 students would be able to test on a computer/device without cards (except for graphs).

Teacher 4 did not want too much to be on the screen – that would be distracting for the student. Teacher 4 did like that the test was on the computer.

Teacher 4 would prefer to have a 1 pager for directions.

Suggestions from Teacher 4:

- Need to know about what can be used – smartboards, iPads, and what needs to happen to use them (rules around security)
- Cards would require two people for administration in current configuration
- Teachers need help understand they can use all of the technology that they use in instruction.
- Cards should be numbered so they can be put back in order.
- Would like a tutorial with questions that will prepare the student for this assessment.
- Limit directions on screen or make sure that they know what to read.
- Progress indicator – could be numbered questions (1 of 15) so that Assessor can let the student know when getting close to the end. (Tend to lose track of which question are on when testing)

Suggestions, comments and notes from DPI and TOPS after watching this administration:

- Need rules for projecting content onto smartboards or other devices.
- Tutorial/demo to walk through the system and to prepare the students

Cleveland HS Observation December 4, 2017

Teacher 5, STC and LEA TC were present

There was an issue with the website being blocked, and we were not able to access the web address on a district computer to perform the system trial. After an over an hour of trying to

have IT bypass the filter for the website, we used a phone to access the website and the cards for administration.

During the administration, an error was made in selecting an answer choice. This brought up the idea of a resume -1 (resume to the last question) and who would be able to perform this function, since it would be during the last 10 days of the year.

Due to the system not functioning, the test administrator used the cards, and the online system was viewed from a smartphone. Teacher 5 indicated that the students in current class could use the computer. The students in the other teacher's class might need the cards.

Students only had access to a Chromebook for 45 minutes on Tuesdays.

Teacher 5 indicated that students need the least amount of stimulation on the screen as possible.

Wanted to know how this would work with a student who is visually impaired – especially on the reading assessment, where the items cannot be read to the student.

Suggestions, comments and notes from this administration:

- Resume -1 (resume to the previous question)
 - This would void the answer entered incorrectly
 - The updated answer would need to be entered in the system
 - This would need to be indicated immediately and fixed on the spot, therefore was suggested that a STC be able to perform this function
 - Cleveland DTC agreed that the STC could perform this function during testing
- Tutorial or practice test would be necessary for students

Appendix 2–F

Item Tryout Comments

Item 1

Form1: We understand these items were based on the extended content standards; our students struggle with these concepts in the classroom as well.

Form 2: Student have to be able to operate a calculator and also solve algebraic expressions.

Form3: I find it frustrating with a student I know to be pre-symbolic. He guessed well on the test, but I know he does not understand the material

Form3: Most of my students who are ID moderate or autistic have not worked with exponential numbers. Most have not mastered basic mathematical operations.

Form5: This item was difficult for my student's cognitive ability.

Form6: For lower level students the awareness did not seem to impact them either way. They randomly picked choices. For example one of my students does not yet know numbers 1-5. For students on higher levels, the content frustrated them.

Form6: We are learning how to skip count, add, subtract, and multiply. The concept of squaring is very high.

Form8: I would like to see the questions more relevant to our population. For example, relate the question to a real-world scenario (if possible).

Form8: Students in the self-contained setting display difficulties comprehending materials on this level.

Form9: Students functional levels are working on number identification, addition, subtraction. Higher functioning students are working on multiplication and division - Not a functional problem - needs to be a word problem with a formula. Is not broken down enough

Form10: This test was really bit to high academically for my students. I notice that for a couple of them the calculator was a distraction and none of my students used it at all.

Item 2

Form1: We understand these items were based on the extended content standards; our students struggle with these concepts in the classroom as well.

Form5: Our students are at a very basic level of learning, algebraic expressions and other similar applications are way too much for their comprehension levels.

Form6: For lower level students the awareness did not seem to impact them either way. They randomly picked choices. For example one of my students does not yet know numbers 1-5. For students on higher levels, the content frustrated them.

Form8: Test items were too difficult for students to comprehend.

Form9: not a functional skill for this population

Item 3

Form1: We understand these items were based on the extended content standards; our students struggle with these concepts in the classroom as well.

Form5: More appropriate for our students

Form8: Question is too complex for students to comprehend the equation.

Form9: We found the question to be very confusing. It was difficult to determine a correct answer when we could not figure out how to interpret the question in the first place

Form10: This test was really bit to high academically for my students. I notice that for a couple of them the calculator was a distraction and none of my students used it at all.

Form10: To hard for non-diploma track students.

Item 4

Form1: We understand these items were based on the extended content standards; our students struggle with these concepts in the classroom as well.

Form5: Slope has not been touched much in our curriculum

Form6: Best question for higher level students.

Form6: This is extremely relatable to all students due to being centered around going to the movies

Form8: An appropriate question

Form8: Great question for this population.

Form8: This test item is more aligned with the needs of student in this population. The graphics were simplistic, and students had the ability to comprehend the question.

Form10: This test was really bit to high academically for my students. I notice that for a couple of them the calculator was a distraction and none of my students used it at all.

Item 5

Form5: This item is very unacceptable, this type of question is setting our students up for failure.

Form7: Smaller number patterns are better as many students can't conceive of numbers above 10

Form8: Good question for this population

Form8: Students do not have the cognitive ability to comprehend questions on this level.

Form10: This test was really bit to high academically for my students. I notice that for a couple of them the calculator was a distraction and none of my students used it at all.

Item 6

Form1: Appeared to be too advance for our population.

Form1: We understand these items were based on the extended content standards; our students struggle with these concepts in the classroom as well.

Form7: Students have minimal understanding of slope and negative numbers

Form8: Students do not have the cognitive ability to comprehend questions on this level.

Form10: This item needs to be corrected - (3,4) and (1,12) do not go together as a function with (0,0)

Form10: This test was really bit to high academically for my students. I notice that for a couple of them the calculator was a distraction and none of my students used it at all.

Form10: Vocabulary is to hard for non-diploma track students.

Item 7

Form 1: We understand these items were based on the extended content standards; our students struggle with these concepts in the classroom as well.

Form3: I do not understand the purpose of this question.

Form3: This question was an issue across the board. Students did not understand at all.

Form7: Students will not understand variables

Form8: Students do not have the cognitive ability to comprehend questions on this level.

Form9: If teaching this skill this not teaching functional skills needed by students

Form10: Students at this level have no idea what slope is.

Form10: This test was really bit to high academically for my students. I notice that for a couple of them the calculator was a distraction and none of my students used it at all.

Item 8

Form 1: #8 appeared to be too advance for our population.

Form1: We understand these items were based on the extended content standards; our students struggle with these concepts in the classroom as well.

Form6: This is a great question - More like this one!

Form8: Students do not have the cognitive ability to comprehend questions on this level.

Form10: This test was really bit to high academically for my students. I notice that for a couple of them the calculator was a distraction and none of my students used it at all.

Item 9

Form2: This style of questioning and wording is most appropriate for mild-moderate.

Form3: Most of the questions are too wordy

Form5: Ridiculous question for extend 1 students.

Form8: Students do not have the cognitive ability to comprehend questions on this level.

Form10: This test was really bit to high academically for my students. I notice that for a couple of them the calculator was a distraction and none of my students used it at all.

Item 10

Form1: These test questions were very difficult for this level of student.

Form1: We understand these items were based on the extended content standards; our students struggle with these concepts in the classroom as well. We would like to see more incorporation of life skills into their assessments, perhaps a question using money or time would be beneficial.

Form5: More appropriate than last question.

Form8: Students do not have the cognitive ability to comprehend questions on this level.

Form10: Students at this level have no idea what an outlier is in the problem.

Form10: This test was really bit to high academically for my students. I notice that for a couple of them the calculator was a distraction and none of my students used it at all.

Appendix 3–A

Fairness and DIF Review Process

Appendix 5–A

Testing Students with disability

<https://www.dpi.nc.gov/media/9770/open>

Appendix 5–B

Testing Security Protocols and Procedures for School Personnel

<https://www.dpi.nc.gov/media/118/open>

Appendix 5–C

North Carolina Test Coordinators’ Policies and Procedures Handbook

<https://www.dpi.nc.gov/media/8677/open>

Appendix 5–D

North Carolina Alternate Assessment Decision Making Flow Chart

[north-carolina-alternate-assessment-final-decision.pdf \(ncpublicschools.gov\)](#)

Appendix 5-E

Online Testing Irregularity Submission System (OTISS) User Manual

https://files.nc.gov/dpi/documents/files/otiss_user_manual.pdf

Appendix 7–A

**Standard Setting Technical Report: Edition 4 NCEXTEND1 Mathematics
2019**

[Technical Information for State Tests | NC DPI](#)

Appendix 7–B

**Observation Report: Edition 4 NCEXTEND1 Mathematics Standard Setting
2018-19**

Appendix 8–A

Subgroups Distribution: NCEXTEND1 Mathematics Scale Score Descriptive Statistics

Table 1. 2018-19 NCEXTEND1 Grade 3 Scale Score Descriptive Statistics by Subgroups

Group	Categories	N	Statistics		Range		Percentile		
			Average	SD	Min	Max	25th	Median	75th
Ethnicity	Asian	54	449	9	419	469	445	449	454
	Black	427	451	9	419	469	446	450	456
	Hispanic	240	450	8	419	470	445	449	455
	American Indian	11	451	9	433	467	444	450	454
	Multiracial	59	451	9	421	466	445	449	459
	Native Hawaiian/Pacific Islander*/	4	449	1	447	449	448	449	449
	White	480	450	9	419	469	445	450	456
SWD	Autism	539	451	9	419	469	446	450	457
	Deaf-Blindness	2	453	4	450	455	450	453	455
	Serious Emotional Disability	2	459	2	457	460	457	459	460
	Hearing Impairment	2	464	2	462	465	462	464	465
	ID	468	450	8	421	470	445	449	455
	Specific Learning Disability	12	454	5	447	464	452	454	457
	Multiple Disability	163	445	10	419	468	442	446	451
	Other Health Impairment	43	452	11	419	468	449	452	457
	Orthopedic Impairment	5	451	8	440	461	449	449	455
	Other	26	450	9	419	468	447	449	453
	Traumatic Brain Injury	12	447	11	421	467	443	448	452
	Visual Impairment	1	436		436	436	436	436	436
EDS	Not Economically Disadvantaged	506	449	9	419	469	445	449	455
	Economically Disadvantaged	769	450	9	419	470	446	449	456
ELs	Regular	1,099	450	9	419	470	445	449	456
	Other	3	447	6	441	452	441	449	452
	English Language Learner	173	450	8	419	468	446	449	454
	All	1,275	450	9	419	470	445	449	455

¹Not Proficient=Inconsistent Understanding, 3= Sufficient Understanding, 4= Thorough Understanding, on Track for Competitive Employment

Table 2. 2018-19 NCEXTEND1 Grade 4 Scale Score Descriptive Statistics by Subgroups

Group	Categories	N	Statistics		Range		Percentile		
			Average	SD	Min	Max	25th	Median	75th
Ethnicity	Asian	44	450	8	421	465	445	449	457
	Black	464	450	9	419	473	445	449	455
	Hispanic	242	451	9	419	474	446	450	457
	American Indian	25	451	13	419	471	447	453	458
	Multiracial	66	449	11	419	472	445	449	455
	Native Hawaiian/Pacific Islander*/	2	454	7	449	459	449	454	459
	White	469	450	9	419	472	445	449	455
SWD	Autism	535	451	9	419	474	446	449	456
	Deaf-Blindness	2	439	25	421	456	421	439	456
	Deafness	2	443	5	439	446	439	443	446
	Serious Emotional Disability	6	455	9	447	469	448	453	462
	Hearing Impairment	1	443		443	443	443	443	443
	ID	509	450	8	419	473	445	449	455
	Specific Learning Disability	5	457	4	451	463	455	456	459
	Multiple Disability	154	447	11	419	471	443	448	453
	Other Health Impairment	51	452	11	419	474	448	453	457
	Orthopedic Impairment	5	448	5	442	454	444	447	451
	Other	24	447	13	419	472	441	447	456
	Traumatic Brain Injury	16	447	10	420	459	444	448	454
	Visual Impairment	2	456	6	451	460	451	456	460
EDS	Not Economically Disadvantaged	527	450	9	419	473	445	449	455
	Economically Disadvantaged	785	450	9	419	474	445	449	456
ELs	Regular	1,128	450	9	419	474	445	449	455
	Other	6	445	5	441	455	441	443	445
	English Language Learner	178	452	9	421	474	446	451	458
	All	1,312	450	9	419	474	445	449	455

¹Not Proficient=Inconsistent Understanding, 3= Sufficient Understanding, 4= Thorough Understanding, on Track for Competitive Employment

Table 3. 2018-19 NCEXTEND1 Grade 5 Scale Score Descriptive Statistics by Subgroups

Group	Categories	N	Statistics		Range		Percentile		
			Average	SD	Min	Max	25th	Median	75th
Ethnicity	Asian	32	448	10	432	470	440	448	454
	Black	468	451	9	417	474	445	450	456
	Hispanic	228	450	8	417	473	445	449	453
	American Indian	10	451	7	440	468	447	451	455
	Multiracial	66	449	9	428	471	444	449	453
	Native Hawaiian/Pacific Islander*/	3	441	6	436	447	436	439	447
	White	518	450	9	418	474	445	449	454
SWD	Autism	493	450	9	419	474	445	449	455
	Deafness	1	461		461	461	461	461	461
	Serious Emotional Disability	5	458	6	453	468	456	456	457
	Hearing Impairment	4	457	12	447	470	447	455	467
	ID	555	450	9	417	474	445	449	454
	Specific Learning Disability	10	467	4	461	474	464	467	470
	Multiple Disability	142	447	10	417	471	442	447	452
	Other Health Impairment	69	452	10	419	474	445	450	457
	Orthopedic Impairment	4	452	7	445	462	447	450	457
	Other	28	446	8	432	466	441	446	451
	Speech or Language Impairment	1	446		446	446	446	446	446
	Traumatic Brain Injury	13	449	10	439	466	442	445	454
EDS	Not Economically Disadvantaged	543	449	9	417	474	445	449	454
	Economically Disadvantaged	782	450	9	417	474	445	449	455
ELs	Regular	1,172	450	9	417	474	445	449	455
	Other	4	449	3	445	453	447	450	452
	English Language Learner	149	450	8	419	473	445	450	454
	All	1,325	450	9	417	474	445	449	454

¹Not Proficient=Inconsistent Understanding, 3= Sufficient Understanding, 4= Thorough Understanding, on Track for Competitive Employment

Table 4. 2018-19 NCEXTEND1 Grade 6 Scale Score Descriptive Statistics by Subgroups

Group	Categories	N	Statistics		Range		Percentile		
			Average	SD	Min	Max	25th	Median	75th
Ethnicity	Asian	51	445	11	419	462	439	448	453
	Black	473	451	9	419	473	446	451	456
	Hispanic	205	450	8	419	470	445	450	455
	American Indian	14	459	9	443	470	452	463	466
	Multiracial	64	451	10	419	473	444	451	456
	Native Hawaiian/Pacific Islander*/	2	447	6	443	451	443	447	451
	White	523	450	9	419	472	445	451	456
SWD	Autism	478	451	9	419	472	445	452	456
	Deaf-Blindness	2	441	3	439	443	439	441	443
	Deafness	2	447	1	446	448	446	447	448
	Serious Emotional Disability	4	461	9	449	469	456	464	467
	Hearing Impairment	3	447	7	439	453	439	450	453
	ID	563	451	8	419	473	446	451	456
	Specific Learning Disability	5	458	5	454	464	454	456	462
	Multiple Disability	166	445	11	419	469	441	447	451
	Other Health Impairment	66	453	7	439	472	447	453	457
	Orthopedic Impairment	1	452		452	452	452	452	452
	Other	18	449	6	435	457	445	450	454
	Speech or Language Impairment	2	450	4	447	453	447	450	453
	Traumatic Brain Injury	18	452	9	434	464	446	453	458
	Visual Impairment	4	458	9	450	470	452	455	463
EDS	Not Economically Disadvantaged	530	449	9	419	472	445	450	454
	Economically Disadvantaged	802	451	9	419	473	446	452	456
ELs	Regular	1,198	450	9	419	473	445	451	456
	Other	7	447	13	419	458	444	452	456
	English Language Learner	127	450	8	419	468	445	451	455
	All	1,332	450	9	419	473	445	451	456

¹Not Proficient=Inconsistent Understanding, 3= Sufficient Understanding, 4= Thorough Understanding, on Track for Competitive Employment

Table 5. 2018-19 NCEXTEND1 Grade 7 Scale Score Descriptive Statistics by Subgroups

Group	Categories	N	Statistics		Range		Percentile		
			Average	SD	Min	Max	25th	Median	75th
Ethnicity	Asian	29	448	15	414	475	444	449	456
	Black	462	450	9	414	476	444	449	455
	Hispanic	211	450	8	429	471	444	449	454
	American Indian	15	457	9	443	472	450	457	464
	Multiracial	57	451	8	441	471	445	448	454
	Native Hawaiian/Pacific Islander*/	1	462		462	462	462	462	462
	White	592	450	9	414	477	444	449	455
SWD	Autism	466	451	9	414	477	445	449	455
	Deaf-Blindness	1	449		449	449	449	449	449
	Deafness	2	451	9	444	457	444	451	457
	Serious Emotional Disability	5	458	8	450	471	454	456	458
	Hearing Impairment	2	449	3	447	451	447	449	451
	ID	622	450	9	414	476	445	449	455
	Specific Learning Disability	3	449	1	448	450	448	449	450
	Multiple Disability	173	446	10	414	471	443	447	451
	Other Health Impairment	51	452	10	431	471	443	451	458
	Orthopedic Impairment	6	449	9	437	464	444	447	452
	Other	12	451	8	436	464	447	452	456
	Speech or Language Impairment	1	461		461	461	461	461	461
	Traumatic Brain Injury	19	450	8	440	472	442	448	457
	Visual Impairment	4	467	5	464	474	464	465	470
EDS	Not Economically Disadvantaged	490	450	9	414	477	444	449	455
	Economically Disadvantaged	877	450	9	414	476	445	449	455
ELs	Regular	1,230	450	9	414	477	444	449	455
	Other	1	441		441	441	441	441	441
	English Language Learner	136	450	9	414	471	444	449	454
	All	1,367	450	9	414	477	444	449	455

¹Not Proficient=Inconsistent Understanding, 3= Sufficient Understanding, 4= Thorough Understanding, on Track for Competitive Employment

Table 6. 2018-19 NCEXTEND1 Grade 8 Scale Score Descriptive Statistics by Subgroups

Group	Categories	N	Statistics		Range		Percentile		
			Average	SD	Min	Max	25th	Median	75th
Ethnicity	Asian	26	448	11	420	468	445	449	453
	Black	440	450	9	418	473	445	449	455
	Hispanic	193	449	9	418	471	444	448	454
	American Indian	15	454	8	440	469	446	454	460
	Multiracial	49	450	9	435	471	444	448	457
	White	547	449	9	418	473	445	449	454
SWD	Autism	426	450	9	418	472	445	448	455
	Deaf-Blindness	1	463		463	463	463	463	463
	Serious Emotional Disability	4	457	9	450	468	450	454	463
	ID	594	451	8	418	473	446	450	455
	Specific Learning Disability	7	458	6	449	468	454	456	463
	Multiple Disability	157	445	10	418	469	442	446	450
	Other Health Impairment	51	451	10	418	473	446	450	456
	Orthopedic Impairment	1	461		461	461	461	461	461
	Other	12	449	7	441	464	444	447	455
	Traumatic Brain Injury	17	451	11	418	471	448	452	457
EDS	Not Economically Disadvantaged	512	449	9	418	473	445	448	454
	Economically Disadvantaged	758	450	9	418	473	445	449	455
ELs	Regular	1,137	450	9	418	473	445	448	454
	Other	5	446	3	443	450	446	446	447
	English Language Learner	128	451	9	418	471	445	450	456
	All	1,270	450	9	418	473	445	449	455

¹Not Proficient=Inconsistent Understanding, 3= Sufficient Understanding, 4= Thorough Understanding, on Track for Competitive Employment

Table 7. 2018-19 NCEXTEND1 Grade 10 Scale Score Descriptive Statistics by Subgroups

Group	Categories	N	Statistics		Range		Percentile		
			Average	SD	Min	Max	25th	Median	75th
Ethnicity	Asian	28	448	10	418	462	444	448	455
	Black	352	451	8	414	470	446	450	455
	Hispanic	130	449	9	414	469	446	451	454
	American Indian	17	453	6	443	464	450	454	457
	Multiracial	45	452	8	436	467	446	451	459
	White	487	450	10	414	472	445	450	457
SWD	Autism	354	450	9	414	470	445	449	455
	Deaf-Blindness	2	455	3	453	457	453	455	457
	Deafness	1	460		460	460	460	460	460
	Serious Emotional Disability	2	450	1	449	450	449	450	450
	Hearing Impairment	4	459	7	455	470	455	456	464
	ID	486	451	8	414	472	447	451	457
	Specific Learning Disability	4	458	3	453	461	456	458	460
	Multiple Disability	139	446	11	414	467	441	447	452
	Other Health Impairment	35	454	9	423	471	450	453	459
	Orthopedic Impairment	8	446	8	434	459	443	445	452
	Other	12	449	13	414	460	446	453	458
	Traumatic Brain Injury	10	453	5	446	460	450	454	458
	Visual Impairment	2	456	11	448	464	448	456	464
EDS	Not Economically Disadvantaged	454	449	9	414	472	445	450	455
	Economically Disadvantaged	605	451	9	414	471	446	451	455
ELs	Regular	955	450	9	414	472	446	450	455
	O	1	462		462	462	462	462	462
	English Language Learner	103	450	9	416	469	446	450	455
All		1,059	450	9	414	472	446	450	455

¹Not Proficient=Inconsistent Understanding, 3= Sufficient Understanding, 4= Thorough Understanding, on Track for Competitive Employment

Appendix 8–B

Achievement Level Ranges and Descriptors

<https://www.dpi.nc.gov/media/9642/open>

Appendix 8–C

NCEXTEND1 Proficiency Classifications by Subgroups

Table 1. 2018-19 NCEXTEND1 Grade 3 Proficiency Classifications by Subgroups

Category	Sub-Category	N	1) Not Proficient, Not CCR	2) Sufficient Understanding, Not CCR	3) Thorough Understanding, CCR
Ethnicity	Asian	54	61.1	33.3	5.6
	Black	427	55.0	34.9	10.1
	Hispanic	240	60.4	33.8	5.8
	American Indian	11	54.6	27.3	18.2
	Multiracial	59	52.5	42.4	5.1
	Native Hawaiian/Pacific Islander*/	4	100.0		
	White	480	54.4	41.5	4.2
SWD	Autism	539	51.8	39.2	9.1
	Deaf-Blindness	2	50.0	50.0	
	Serious Emotional Disability	2		100.0	
	Hearing Impairment	2		50.0	50.0
	ID	468	56.6	37.8	5.6
	Specific Learning Disability	12	25.0	66.7	8.3
	Multiple Disability	163	74.9	24.5	0.6
	Other Health Impairment	43	37.2	51.2	11.6
	Orthopedic Impairment	5	60.0	40.0	
	Other	26	61.5	34.6	3.9
	Traumatic Brain Injury	12	75.0	16.7	8.3
	Visual Impairment	1	100.0		
EDS	Not Economically Disadvantaged	506	57.3	36.6	6.1
	Economically Disadvantaged	769	55.3	37.7	7.0
ELs	Regular	1,099	55.2	38.1	6.6
	Other	3	66.7	33.3	
	English Language Learner	173	61.3	31.8	6.9
	All	1,275	56.1	37.3	6.7

¹Not Proficient=Inconsistent Understanding, 3= Sufficient Understanding, 4= Thorough Understanding, on Track for Competitive Employment

Table 2. 2018-19 NCEXTEND1 Grade 4 Proficiency Classifications by Subgroups

Category	Sub-Category	N	1) Not Proficient, Not CCR	2) Sufficient Understanding, Not CCR	3) Thorough Understanding, CCR
Ethnicity	Asian	44	56.8	40.9	2.3
	Black	464	61.4	32.1	6.5
	Hispanic	242	52.5	40.5	7.0
	American Indian	25	44.0	40.0	16.0
	Multiracial	66	54.6	39.4	6.1
	Native Hawaiian/Pacific Islander*/	2	50.0	50.0	
	White	469	60.1	34.1	5.8
SWD	Autism	535	55.5	37.9	6.5
	Deaf-Blindness	2	50.0	50.0	
	Deafness	2	100.0		
	Serious Emotional Disability	6	50.0	33.3	16.7
	Hearing Impairment	1	100.0		
	ID	509	61.5	32.0	6.5
	Specific Learning Disability	5		100.0	
	Multiple Disability	154	64.9	30.5	4.6
	Other Health Impairment	51	43.1	45.1	11.8
	Orthopedic Impairment	5	60.0	40.0	
	Other	24	58.3	37.5	4.2
	Traumatic Brain Injury	16	68.8	31.3	
	Visual Impairment	2		100.0	
EDS	Not Economically Disadvantaged	527	62.2	31.5	6.3
	Economically Disadvantaged	785	55.9	37.7	6.4
ELs	Regular	1,128	60.2	33.6	6.2
	Other	6	83.3	16.7	
	English Language Learner	178	46.6	46.1	7.3
	All	1,312	58.5	35.2	6.3

¹Not Proficient=Inconsistent Understanding, 3= Sufficient Understanding, 4= Thorough Understanding, on Track for Competitive Employment

Table 3. 2018-19 NCEXTEND1 Grade 5 Proficiency Classifications by Subgroups

Category	Sub-Category	N	1) Not Proficient, Not CCR	2) Sufficient Understanding, Not CCR	3) Thorough Understanding, CCR
Ethnicity	Asian	32	68.8	25.0	6.3
	Black	468	60.5	29.5	10.0
	Hispanic	228	64.0	30.3	5.7
	American Indian	10	70.0	20.0	10.0
	Multiracial	66	71.2	22.7	6.1
	Native Hawaiian/Pacific Islander*/	3	100.0		
	White	518	64.7	28.6	6.8
SWD	Autism	493	64.3	28.6	7.1
	Deafness	1		100.0	
	Serious Emotional Disability	5		80.0	20.0
	Hearing Impairment	4	50.0	25.0	25.0
	ID	555	62.5	29.4	8.1
	Specific Learning Disability	10		30.0	70.0
	Multiple Disability	142	72.5	26.1	1.4
	Other Health Impairment	69	56.5	30.4	13.0
	Orthopedic Impairment	4	75.0	25.0	
	Other	28	78.6	17.9	3.6
	Speech or Language Impairment	1	100.0		
	Traumatic Brain Injury	13	69.2	23.1	7.7
EDS	Not Economically Disadvantaged	543	66.7	26.3	7.0
	Economically Disadvantaged	782	61.5	30.3	8.2
ELs	Regular	1,172	63.9	28.0	8.1
	Other	4	75.0	25.0	
	English Language Learner	149	61.1	34.2	4.7
	All	1,325	63.6	28.7	7.7

¹Not Proficient=Inconsistent Understanding, 3= Sufficient Understanding, 4= Thorough Understanding, on Track for Competitive Employment

Table 4. 2018-19 NCEXTEND1 Grade 6 Proficiency Classifications by Subgroups

Category	Sub-Category	N	1) Not Proficient, Not CCR	2) Sufficient Understanding, Not CCR	3) Thorough Understanding, CCR
Ethnicity	Asian	51	72.6	27.5	
	Black	473	57.9	37.4	4.7
	Hispanic	205	58.5	37.1	4.4
	American Indian	14	28.6	28.6	42.9
	Multiracial	64	57.8	32.8	9.4
	Native Hawaiian/Pacific Islander*/	2	100.0		
	White	523	57.9	36.0	6.1
SWD	Autism	478	55.2	38.7	6.1
	Deaf-Blindness	2	100.0		
	Deafness	2	100.0		
	Serious Emotional Disability	4	25.0	25.0	50.0
	Hearing Impairment	3	66.7	33.3	
	ID	563	56.5	37.1	6.4
	Specific Learning Disability	5		80.0	20.0
	Multiple Disability	166	79.5	19.3	1.2
	Other Health Impairment	66	48.5	47.0	4.6
	Orthopedic Impairment	1	100.0		
	Other	18	66.7	33.3	
	Speech or Language Impairment	2	50.0	50.0	
	Traumatic Brain Injury	18	50.0	44.4	5.6
	Visual Impairment	4	25.0	50.0	25.0
EDS	Not Economically Disadvantaged	530	63.2	32.3	4.5
	Economically Disadvantaged	802	55.1	38.5	6.4
ELs	Regular	1,198	58.4	35.6	6.0
	Other	7	57.1	42.9	
	English Language Learner	127	57.5	40.2	2.4
	All	1,332	58.3	36.0	5.6

¹Not Proficient=Inconsistent Understanding, 3= Sufficient Understanding, 4= Thorough Understanding, on Track for Competitive Employment

Table 5. 2018-19 NCEXTEND1 Grade 7 Proficiency Classifications by Subgroups

Category	Sub-Category	N	1) Not Proficient, Not CCR	2) Sufficient Understanding, Not CCR	3) Thorough Understanding, CCR
Ethnicity	Asian	29	55.2	34.5	10.3
	Black	462	52.4	42.6	5.0
	Hispanic	211	55.9	39.3	4.7
	American Indian	15	13.3	66.7	20.0
	Multiracial	57	54.4	38.6	7.0
	Native Hawaiian/Pacific Islander*/	1		100.0	
	White	592	52.2	42.9	4.9
SWD	Autism	466	50.4	44.2	5.4
	Deaf-Blindness	1	100.0		
	Deafness	2	50.0	50.0	
	Serious Emotional Disability	5		80.0	20.0
	Hearing Impairment	2	50.0	50.0	
	ID	622	51.9	42.3	5.8
	Specific Learning Disability	3	66.7	33.3	
	Multiple Disability	173	64.7	33.0	2.3
	Other Health Impairment	51	47.1	45.1	7.8
	Orthopedic Impairment	6	66.7	33.3	
	Other	12	25.0	75.0	
	Speech or Language Impairment	1		100.0	
	Traumatic Brain Injury	19	63.2	31.6	5.3
	Visual Impairment	4		75.0	25.0
EDS	Not Economically Disadvantaged	490	53.7	41.4	4.9
	Economically Disadvantaged	877	51.9	42.7	5.5
ELs	Regular	1,230	52.3	42.5	5.2
	Other	1	100.0		
	English Language Learner	136	54.4	39.7	5.9
	All	1,367	52.5	42.2	5.3

¹Not Proficient=Inconsistent Understanding, 3= Sufficient Understanding, 4= Thorough Understanding, on Track for Competitive Employment

Table 6. 2018-19 NCEXTEND1 Grade 8 Proficiency Classifications by Subgroups

Category	Sub-Category	N	1) Not Proficient, Not CCR	2) Sufficient Understanding, Not CCR	3) Thorough Understanding, CCR
Ethnicity	Asian	26	73.1	23.1	3.9
	Black	440	64.6	28.4	7.1
	Hispanic	193	71.0	21.8	7.3
	American Indian	15	46.7	46.7	6.7
	Multiracial	49	63.3	30.6	6.1
	White	547	68.7	24.5	6.8
SWD	Autism	426	67.8	23.2	8.9
	Deaf-Blindness	1		100.0	
	Serious Emotional Disability	4	50.0	25.0	25.0
	ID	594	64.0	29.6	6.4
	Specific Learning Disability	7	14.3	71.4	14.3
	Multiple Disability	157	85.4	12.1	2.6
	Other Health Impairment	51	58.8	33.3	7.8
	Orthopedic Impairment	1		100.0	
	Other	12	66.7	33.3	
	Traumatic Brain Injury	17	58.8	35.3	5.9
EDS	Not Economically Disadvantaged	512	69.3	24.8	5.9
	Economically Disadvantaged	758	65.8	26.7	7.5
ELs	Regular	1,137	67.4	26.2	6.4
	Other	5	100.0		
	English Language Learner	128	64.8	24.2	10.9
	All	1,270	67.2	25.9	6.9

¹Not Proficient=Inconsistent Understanding, 3= Sufficient Understanding, 4= Thorough Understanding, on Track for Competitive Employment

Table 7. 2018-19 NCEXTEND1 Grade 10 Proficiency Classifications by Subgroups

Category	Sub-Category	N	1) Not Proficient, Not CCR	2) Sufficient Understanding, Not CCR	3) Thorough Understanding, CCR
Ethnicity	Asian	28	64.3	35.7	
	Black	352	57.1	37.8	5.1
	Hispanic	130	54.6	42.3	3.1
	American Indian	17	41.2	47.1	11.8
	Multiracial	45	55.6	31.1	13.3
	White	487	57.1	37.0	6.0
SWD	Autism	354	62.2	31.9	5.9
	Deaf-Blindness	2		100.0	
	Deafness	1		100.0	
	Serious Emotional Disability	2	100.0		
	Hearing Impairment	4		75.0	25.0
	ID	486	50.8	43.2	6.0
	Specific Learning Disability	4		100.0	
	Multiple Disability	139	73.4	24.5	2.2
	Other Health Impairment	35	34.3	54.3	11.4
	Orthopedic Impairment	8	75.0	25.0	
	Other	12	50.0	50.0	
	Traumatic Brain Injury	10	40.0	60.0	
	Visual Impairment	2	50.0		50.0
EDS	Not Economically Disadvantaged	454	60.8	33.9	5.3
	Economically Disadvantaged	605	53.6	40.7	5.8
ELs	Regular	955	57.1	37.0	6.0
	O	1		100.0	
	English Language Learner	103	53.4	44.7	1.9
	All	1,059	56.7	37.8	5.6

¹Not Proficient=Inconsistent Understanding, 3= Sufficient Understanding, 4= Thorough Understanding, on Track for Competitive Employment

Appendix 8–D

Score Interpretive guide

[Interpretive Guide to the North Carolina Winscan Score Reports \(nc.gov\)](#)