The North Carolina Department of Public Instruction
NCEXTEND1 Alternate Assessments
Edition 4 Technical Report
Reading Grades 3–8, and English II at Grade 10
Science Grades 5, 8, and Biology at Grade 10
2020–2021

Prepared by:
North Carolina Department of Public Instruction
Accountability Division
2022
Tammy Howard, Ph.D., Director of Accountability
Kinge Mbella, Ph.D., Lead Psychometrician
Thakur Karkee, Ph. D., Psychometrician
Maxey-Moore, Section Chief, Test Development

Table of Contents

Table of Tables

Table of Figures

CHAPTER 1 INTRODUCTION

The intent of this technical report is to provide comprehensive and detailed description of all steps implemented towards the development, analysis, and reporting of test scores from the NCEXTEND1 tests. Technical evidence presented throughout this report also serve as primary sources of validity to support intended test score uses and interpretation. The validity evidence is documented in terms of processes used in review, revision, and implementation of new content standards; develop test specifications and items; field-test and item analysis; bias and sensitivity review; test development; scoring and scale development; and standard setting.

The first part of this report presents a brief overview of the revision and eventual adoption of new Reading at grades 3–8 and English II at grade 10 extended content standards (ECS) which is bases for the development of new assessments. Subsequently, processes are briefly described for 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 of summative assessments for the *Edition 4* Reading at grades 3–8 and 10.

The NCDPI recommends interpreting 2020–21 summary results cautiously as circumstances of the school year were affected by COVID-19 pandemic. First, COVID-19 related disruptions to normal learning and school environments lead to varied instructional practices across public school systems (PSU) in the state including in-person, virtual, and hybrid instructions. Second, in 2020-21 school year the United States Department of Education and State of North Carolina waived accountability which implied the high stakes consequences usually attributed to test scores did not apply in 2020-21. Finally, the accountability waiver also applied to the 95% participation requirement. Even though participation rate for state assessments in 2020-21 were close to expected 95%, however, participation rates across districts and subgroups varied and there is no direct evidence that the missingness was random. As a result of these circumstances, caution is advised when attempting to compare student performance from 2020-21 with other years.

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. In North Carolina, a student with the most significant cognitive disability is defined as a student:

- whose disability significantly impacts cognitive functioning and adaptive behaviors, defined as those skills which are essential for someone to live and function independently,
- who requires extensive and repeated individualized instruction and support to make meaningful gains, and
- who uses substantially adapted materials and individualized methods of accessing information in alternative ways.

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 high school courses shall participate in the End-Of-Grade (EOG) and End-Of-Course (EOC) state assessment programs adopted by the NCSBE. To participate in the alternate assessment, students must meet eligibility criteria established by the NCDPI. This policy is in accordance with IDEA and ESSA. The 16 N.C. Admin. Code 06G .0315 requires that all students entitled to testing accommodations shall participate in the North Carolina State Annual Testing Program (NCSATP) using one of the following assessments as required by the student's accommodation: (1) The standard test administration with or without accommodations, or (2) An alternate assessment with or without accommodations.

Students entitled to testing accommodations included (1) students with Individualized Education Programs (IEPs) created under the Individuals with Disabilities Education Act, 33 U.S.C. 1414, and regulations adopted pursuant to that Act; (2) students with a plan created under 504 of the Rehabilitation Act of 1973, 29 U.S.C. 794 as implemented through 34 C.F.R. 104.44, and other regulations adopted pursuant to that Act; (3) students with documented transitory impairments with actual or expected duration of six months or less that affect their ability to demonstrate their knowledge on standard test administrations without accommodation as determined by the LEA; and (4) students who score below Level 5.0 Bridging on the reading domain of the WIDA Screener/ACCESS for ELLs®.

The Reading at grades 3–8 and 10 aligned to the North Carolina ECSs for reading and science at grades 5 and 8 and Biology at grade 10 aligned to ECSs for science administered by the North Carolina Department of Public Instruction (NCDPI) measure students' proficiency on the ECSs. The assessment results are used for school and district accountability under the accountability model and for federal reporting purposes.

The North Carolina reading content standards were revised and adopted in 2017. Consequently, NCEXTEND1 alternate assessments were also revised and redesigned as *Edition 4* NCEXTEND1 alternate assessments in 2017. For science only, the design of the alternate assessments were revised to improve the technical quality of these assessments and aligned them with the other alternate assessments in mathematics and reading.

The plan was to conduct field test in 2018–19 administration with implementation of *Edition 4* alternate operational assessments and standards setting in 2019–20. Due to COVID-19 pandemic, all assessments were suspended in 2019–20. However, administration of statewide assessments resumed in 2020–21. However, the NCDPI received waiver from USED for accountability as schools were still dealing with impact and disruption of COVID-19.

Subsequent plan was to use 2020–21 Classical Test Theory (CTT) statistics from operational administration as final statistics for standard setting. Due to likely unstable statistics from the administration, the NCDPI decided to use CTT statistics from the field-test in 2018–19 and review them again in succeeding administrations.

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 for Reading and Science. Due to the COVID-19 pandemic, the student performance is likely to be affected resulting in less than expected proficiency rates. Therefore, the validity and reliability evidence documented in this report are generally borne from the processes used for review, revision, and implementations of the new ECSs; item development and review; field-test and form development; and test administration, scoring, and reporting.

The first part of this report presents a brief overview of the revision, redesigned, and eventual adoption of the new NCEXTEND1 Reading content standards which are used to justify the development of new assessments. The science standards were not revised but alternate assessments were redesigned. 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 the standard setting workshop used to set achievement levels for reporting and interpreting student results, and validity evidence for the *Edition 4* NCEXTEND1 Reading and science summative alternate assessments.

1. 1 Purpose and Background of the North Carolina State Testing Program

The purposes of the North Carolina Annual Testing Program, as described in G.S. §115C-174.10, are as follows:

"*(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 Content Standards.

The NCEXTEND1 alternate assessments are teacher-facilitated online assessments that are administered individually to each student. For the Reading and Science NCEXTEND1 alternate assessments, the questions are presented online in two (2) sets; Set 1 and Set 2. 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 1 for students who do not answer enough questions correctly. Students who answer enough questions correctly in Set 1 will continue to Set 2. For Science, the teacher reads all items. For Reading, all selections from Set 1 are read by the teacher. For Set 2, students are expected to independently read one operational selection student respond to questions associated with that selection. The NCEXTEND1 Alternate Assessments may be administered over several days or may be completed in one session. The time required by a student to complete the assessment is unique for each student, depending on the student's ability to maintain focus, medical condition, and/or fatigue factor(s).

This technical report documents all steps and processes that were implemented in the redesign, development, administration, scoring, and reporting of results for *Edition 4* of NCEXTEND1 alternate Reading and Science 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 (EC) Division, in collaboration with K–12 Standards, Curriculum and Instruction Division (currently Academic Standards Division), developed and implemented a plan of action and timeline in 2016 to review and revise the Reading ECSs. The Science NCEXTEND1 ECSs were not revised, only tests were redesigned in 2017.

Table 1.1 outlines timelines and brief descriptions of actions that were implemented by the NCDPI during the review, revision, redesign, and implementation of the new alternate assessments that are aligned to the new adopted ECSs from 2017 through 2019. 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, redesign, and implementation of the new alternate assessments. The attributes described in *Table 1.1* are a part of validity evidence to show that North Carolina ECSs are research-based and have adequate rigor and expectations to prepare students with significant cognitive disabilities for post-secondary outcomes.

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

Date	Actions	Descriptions
March–April 2017	Revised Reading at grades 3–8 and 10 ECSs and presented comments to NCSBE.	Revised ECSs aligned to the North Carolina Standard Course of Study (NCSCoS) and presented comments and draft to NCSBE for discussion and actions.
April 21, 2017	Science at grades 5 and 8, Biology at grade 10 alternate assessment Redesign Focus Group Meeting.	The Focus Group reviewed assessable Science ECSs, manipulatives, and content-specific topics.
May 2017	Created professional development resources.	Created professional development resources for Reading and Science to support the revisions and redesign.
June 2017	Extended Content Standards Adopted by NCSBE.	NCSBE approved Reading and Science ECSs for adoption.
August 2017	ECSs implementations.	Districts implemented new reading standards and continued to support schools.
August 2017–May 2018	Extended Content Standards implementation preparation.	Conducted professional development for revised and redesign of NCEXTEND1 assessments, developed resources, and revised all support materials
June–August 2018	Summer professional development delivery.	Organized PD webinars.
2018–19	Item development and field-test.	Items based on new standards developed and field tested with new assessment design.
2019–20	New NCEXTEND1 assessments adopted statewide.	Due to COVID-19 pandemic, operational administration and standard setting were held in 2020–21.
2020–21	Standard Setting and score reporting.	Standard setting conducted in July 2021 and scores reported on new achievement level scale.

1. 3 Overview of the North Carolina Reading and Science Alternate Assessment Program

The NCDPI designs NCEXTEND1 alternate assessments for grades 3–8 and 10 Reading, and Science at grades 5 and 8 and Biology at grade 10. These assessments scores provide valid and reliable information intended to serve two general purposes: measure students' performance and progress as it relates to their proficiency towards grade-level content standards and serve as a quantitative indicator for use in federal and statewide accountability models.

- Measure students' performance and progress: North Carolina NCEXTEND1 assessments are used to measure whether students are performing at a level that indicates they consistently demonstrate mastery of the content standards. These assessments are designed to measure student performance on the full breadth and depth of grade-level content standards. Student performance on the assessments is reported using scale scores grouped into one of three achievement levels (Not Proficient, Level 3, and Level 4).
- Federal and State Accountability Models: The NCEXTEND1 assessments are used, as required by federal and state law, as indicators in the school accountability models. These models are designed to identify schools in need of support. Specifically, these assessment scores are used as measures of proficiency and academic growth as defined using SAS© Education Value-Added Assessment System (EVAAS) under the current accountability systems.

The North Carolina *Testing Code of Ethics* (*Appendix 1–A*) cautions educators to use test scores and reports appropriately only for the intended uses as approved by the NCSBE and for which the 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. Test data help educators understand educational patterns and practices. The use of NCEXTEND1 test scores for purposes other than those intended by the 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 is documented throughout this report. Chapter 1 provides a brief history of testing in North Carolina; the standards review, revision, redesign, and implementation process; and overview of the North Carolina statewide assessment program.

Chapter 2 documents an overview of NCSATP test design, item development process, and field-test plans. The test design involved test specifications meetings to specify test blueprints, test complexities, 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 Test Theory (CTT)statistics. The NCDPI has set internal criteria for filtering out items with less-than-optimal quality. Final sections describe summary of item analysis.

Chapter 4 documents 24–step operational form assembly and review processes including content review, production review, and bias and sensitivity review. Finally, the chapter summarizes the structure of the base forms in terms of item types and text complexities, and documents 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 Classical Test Theory based statistics and scaling. Final section describes score certification process.

Chapter 7 presents a summary of the standard setting study that was conducted in July 2021 after the first operational administration of NCEXTEND1 tests. The chapter is a condensed version of the final report prepared by Data Recognition Corporation (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 2020–21 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 reliability and 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
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
NCSCOS	North Carolina Standard Course of Study

Abbreviations	Full Form
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
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 reading and science 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 10 Reading; and Science at grades 5 and 8, and Biology at grade 10 are standard-based alternate assessments that serve summative purposes. The alternate assessments for Reading were aligned with the new ECSs and blueprints adopted in 2017–18 to ensure adequate validity evidence in support of standard-based interpretation of test scores. The format of these alternate assessments for Reading and Science were also redesigned to ensure better alignment to ECS and improved accessibility for all students. The new assessments designs are guided by the overall test specifications which outline all essential content, cognitive, and psychometric specifications.

The NCDPI recruited North Carolina educators from across the state and conducted two on-site test specification workshops in 2018 for alternate assessments. Educators invited to these meetings represented North Carolina educators from across all geographic regions, demographic subgroups, and experiences. The educators also included experience in teaching Special Education and English Learner students to ensure fairness and accessibility of alternate assessment 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 complexity, test blueprints, and psychometric specifications for *Edition 4* alternate assessments.

2.1.1 Content Blueprint

The on-site test specification workshop was facilitated by the NCDPI Test Development staff and designed to get participants to recommend content blueprints for *Edition 4* NCEXTEND1 assessments. 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 assessments followed by an overview of the new 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 assessments, participants were directed to individually assign 0–10 ratings on a survey 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 reviewed the recommended blueprints for the Reading ECSs to ensure adequate across-grades articulation. The final recommendations shown in *Tables 2.1* were then adopted as *Edition 4* Reading content blueprints for NCEXTEND1 assessments.

Table 2.1 NCEXTEND1 Reading Alternate Assessments Test Blueprint (%)

Domain	Grade/Course		
	3–5	6–8	10
Reading for Literature	38–46	38–46	38–46
Reading for Informational Text	46–54	42–50	42–50
Language	4–12	8–16	8–16

The blueprint for science did not change—only test structures redesigned. *Table 2.2* shows the blueprint for the *Edition 4* science.

Table 2.2 NCEXTEND1 Science Alternate Assessment Test Blueprint (%)

Domain	Grade/Course		
	Grade 5	Grade 8	Biology
Life Science: Structures and Functions of Living Organisms (L1)	24–32	24–32	28–36
Life Science: Ecosystems (L2)	24–32	24–32	64–72
Earth and Environmental Science: Earth Systems, Structures, and Processes (E1)	16–24	24–32	
Physical Science: Forces and Motion (P1)	8–16	4–12	
Physical Science: Matter, Properties, and Change (P2)	8–16	8–16	

2.1.2 Cognitive Complexity

Reading: On Day 2 of the test specification workshop, participants were tasked to evaluate and recommend content cognitive complexity expectation ranges for all assessable standards to guide item and test development. The NCDPI adopted the Norman Webb's Depth of Knowledge (DOK) classification (Hess, 2013) as the basis for evaluating content complexity for NCEXTEND1 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.

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 of the test specification workshop, 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 surveys 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.

In Round 2 participants went back to their table and reviewed their ratings based on the large group discussion. The final recommended DOK classifications from Round 2 were then adopted as the expected content cognitive complexity recommendations for assessed Reading content standards. At the conclusion of the meeting, the NCDPI's Test Development and Academic Standards Division staffs reviewed these recommended classifications to ensure coherent

alignment with grade-level content standards expectations and summarized the data into DOK range specifications for the Reading NCEXTEND1 alternate assessments. The final content cognitive complexity specifications for *Edition 4* NCEXTEND1 Reading tests are shown in *Table 2.3*.

Table 2.3 Proposed Target Reading DOKs Across Grades

Grade/Course	Number of Items	Category (%)	
		DOK 1	DOK 2
3–5	24	38–63	38–63
6–8	24	25–46	54–75
10	24	17–30	71–83

Science: The cognitive rigor for science remained the same from 2012–13 which was based on Revised Bloom Taxonomy (RBT).

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 assessment are North Carolina professional educators from across the state who have current classroom experience. Educators who want to serve as item writers or reviewers are required to successfully complete two consecutive day in-person training sessions. The training includes a general course on item writing guidelines, including lessons on sensitivity and bias concerns, North Carolina's *Testing Code of Ethics*, test security agreement, alternate assessment basics, eligibility criteria and student characteristics, extended content standards, item writing practice, and the online test development system. The course provides an overview of the test development process and the basic rules and structures of item formats used by the North Carolina Annual Testing Program. Item writer and reviewer training incorporates the concept of universal design, diversity, fairness, and accessibility to the content being measured. The NCSU-TOPS recruits educators with experience in ECSs for item writing and/or reviewing.

2.3 Item Development Process

The item development process for the Reading NCEXTEND1 *Edition 4* alternate assessment began after the NCSBE adopted the new North Carolinas ECSs. The same process was used for the science item development. North Carolina test items are written and reviewed by trained North Carolina educators who serve as item writers. Additionally, the NCDPI Test Measurement Specialist (TMS) in partnership with Content Specialists at NCSU-TOPS participated in the item development processes. Ultimately, the NCDPI's TMSs served as the final staff reviewer for all

NCEXTEND1 assessments items. Educators with classroom and grade level ECSs experience across the state are recruited, trained, and awarded contracts to write NCEXTEND1 assessment items. The use of educators with classroom experience from across the state for item writing is evidence of instructional validity pertaining to how well the test items reflect classroom instruction

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's iterative 16–step alternate assessment item development and review process. Full details of the 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. A content specialist as subject matter expert, and a specialist each for the Exceptional Children (EC), English Learner (EL), and Visually Impaired (VI) 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 content specialists review the item's assigned ECS, secondary ECS (if applicable), and key/appropriate foils. The content specialist also reviews and makes the following decisions:

- 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 EC educator, and the other is an educator with experience in general education. The EC reviewer 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 Mode of Test Administration

There are three (3) administration options available for the NCEXTEND1 alternate assessment. 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 a 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 the teacher recording the student's 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 the teacher recording the student's responses on the device.

2.5 NCEXTEND1 Reading and Science Alternate Assessment Standalone Field-Test Design, 2018–19

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 including statistical characteristics, accessibility, fairness, and to provide baseline statistical targets to assemble parallel forms. Given the nature of the NCEXTEND1 student population, standalone field-test administration was planned in 2018–19 administration that offered a flexible opportunity to gather essential item level data. Both NCEXTEND1 Reading and Science items were scaffolded with a second trial to answer the item that was incorrectly answered it in the first time by removing it from the display. Each item is worth 2-point if answered correctly during first trial and 1-point if the correct answer is selected during second trial. The NCEXTEND1 assessments are primarily computer-based administered in an online platform displaying the directions, item stem, and answer choices. The entire tests are also available in the paper book or story book format. Based on individual student need, the use of the paper manipulatives is optional. No new item types were introduced in *Edition 4*.

The 2018–19 design for a test coordinator for discontinuing a student from testing on NCEXTEND1 assessments due to his/her non-response was if the student did not respond first six items or obtained five score points or less in the first set. However, the NCDPI Technical Advisory Committee recommended not to use this rule due to equity issue. This rule will be abandoned from the succeeding administrations.

Reading Standalone Field-Test Design: A summary of the Reading standalone field-test conducted in 2018-19 administration is shown in *Table 2.4*. The design for grades 3, 4, 6, and 7 Reading included four (4) forms with 12 selections, six (6) items in each selection matrix sampled with a total of 72 items at each grade. Grades 5, 8, and English II at grade 10 students were divided into two groups and each group either participated in Reading/English II or Science/Biology NCEXTEND1 field-test. Due to relatively smaller sample size from the split, only three (3) forms were field tested with a total of 54 items for these grades.

Each field-test form consisted of three (3) selections: two on Set 1 and one on Set 2 for a total of 18 items.

- Set 1 consisted of two selections with a total of 12 items (six items per selection). Teachers read both selections and all items to students.
- Set 2 consisted of one selection with six items. Students required to read this selection on their own and teachers read the items.

Table 2.4 NCEXTEND1 Reading Standalone Field-Test Plan, 2018–19

Grade/ Course	Number of Items Per Form				Total			Expected Population
	Set 1*	Set 2*	Total No. of Items	No. of Forms	No. of Selections	Items per Selection	Items	
3, 4, 6, 7	12 (2)	6 (1)	18	4	12	6	72	1000
5, 8 and 10	12 (2)	6 (1)	18	3	9	6	54	500**

*Number in parentheses are numbers of selections

** The student population for grades 5, 8, and 10 were divided between reading and science where half of the students participated in reading field test and the half in science in order to avoid double testing.

Each item is worth 2-point if answered correctly after the trial 1. If a student fails to select the correct answer in the trial 1, the incorrect response they selected is removed, and the item is presented again in trial 2 with remaining response options. If the correct answer is selected in trail 2, the student earns 1-point. If the incorrect response option is selected, the student earns 0-point and continue to the next question.

Science Standalone Field-Test Design: A summary of Science field-test design is shown in *Table 2.5*. The plan was to field-test a total of 60 items in each grade level in Science, and Biology at grade 10 NCEXTEND1. Each field-test form consisted of a total of 20 items:

- Set 1 consisted of a total of 12 items worth maximum of 24 points.
- Set 2 consisted of a total of 8 items worth maximum of 16 points.

Table 2.5 NCEXTEND1 Science Standalone Field-Test Plan, 2018–19

Grade/Course	Number of Items				Total No. of Items	Expected Population
	Set 1	Set 2	Total No. of Items	No. of Forms		
5, 8, and Biology	12	8	20	3	60	500*

* The student population for grades 5, 8, and 10 were divided between reading and science where half of the students participated in reading field test and the half in science in order to avoid double testing.

CHAPTER 3 ITEM ANALYSIS

This chapter summarizes procedures and criteria the 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 used 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 for the NCEXTEND1 assessments primarily because the sample size per item was small, mostly less than 200, and low variation of student ability. 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 NCEXTEND1 tests are divided into two sets with Set 2 being relatively more difficult than Set 1. Students who did not respond, assessors use professional judgement to determine when a student is not productively working through the NCEXTEND1 assessment. In the rare case a student has "no response" recorded for question 1 through question 6 (for the first and second trials), assessors may end the test session on question 7. This will result in the lowest possible score for the student. If a student has responded to any of the six (6) questions for the first or second trials, the student is to continue working until the testing session ends.
All field-test items are classified into one of the three NCDPI item flagging categories (Keep, Reserve, and Weak) based on first trial classical statistics with a goal to rank items in the final item pool for overall statistical quality. Items that do not meet the minimum quality requirements were removed from the field test pool later used in form development.

- 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 the NCDPI policy, whenever possible, all items must first undergo field-test prior to selecting on an operational form. After items are field tested, the first step for the NCEXTEND1 assessments involved conducting a series of CTT analysis to determine if these items meet the minimum psychometric requirements to be considered for further evaluation. The 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.

Table 3. 1 CTT Item Flagging Criteria

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

Table 3.2 and *Table 3.3* show descriptive statistics of p-value and point-biserial from the standalone field-test in 2018–19 for the NCEXTEND1 Reading and Science items respectively. The statistics for the first trial are based on responses when students answered the item for the first time. Valid scores in the first trial are either 2 or 0 and for the second trial are 1 or 0.

The descriptive statistics from *Table 3.2* and *Table 3.3* indicated 90% or more items for NCEXTEND1 Reading at grades 3–8 and 10 and Science at grades 5 and 8, and Biology at grade 10 are classified as meeting the NCDPI's optimal standard 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, Reading, Spring 2019

Grade	Flag	Trial	N*	P–value				Point Biserial			
				Mean	SD	Min	Max	Mean	SD	Min	Max
3	Keep	First	71	0.53	0.11	0.28	0.76	0.49	0.10	0.21	0.67
		Second	71	0.77	0.07	0.59	0.89	0.50	0.09	0.21	0.78
	Weak	First	1	0.36	.	0.36	0.36	0.03	.	0.03	0.03
		Second	1	0.67	.	0.67	0.67	0.26	.	0.26	0.26
4	Keep	First	72	0.52	0.12	0.27	0.76	0.47	0.10	0.23	0.64
		Second	72	0.77	0.09	0.55	0.91	0.46	0.09	0.19	0.61
5	Keep	First	53	0.49	0.11	0.27	0.69	0.45	0.09	0.24	0.62
		Second	53	0.76	0.09	0.50	0.91	0.45	0.12	0.18	0.63
	Reserve	First	1	0.32	.	0.32	0.32	0.19	.	0.19	0.19
		Second	1	0.76	.	0.76	0.76	0.28	.	0.28	0.28
6	Keep	First	70	0.52	0.11	0.22	0.71	0.50	0.10	0.19	0.63
		Second	70	0.77	0.08	0.58	0.92	0.48	0.08	0.24	0.62
	Reserve	First	2	0.48	0.00	0.48	0.48	0.16	0.05	0.12	0.19
		Second	2	0.74	0.03	0.72	0.76	0.37	0.03	0.35	0.39
7	Keep	First	69	0.52	0.11	0.33	0.78	0.46	0.10	0.25	0.64
		Second	69	0.78	0.08	0.61	0.94	0.44	0.10	0.20	0.61
8	Reserve	First	3	0.35	0.03	0.32	0.37	0.16	0.01	0.14	0.17
		Second	3	0.69	0.09	0.59	0.76	0.21	0.07	0.14	0.28
	Keep	First	53	0.50	0.10	0.29	0.70	0.44	0.10	0.21	0.64
		Second	53	0.76	0.08	0.56	0.91	0.45	0.09	0.24	0.65
	Reserve	First	1	0.42	.	0.42	0.42	0.17	.	0.17	0.17
		Second	1	0.62	.	0.62	0.62	0.31	.	0.31	0.31
10	Keep	First	51	0.51	0.09	0.28	0.67	0.46	0.10	0.25	0.65
		Second	51	0.77	0.09	0.51	0.89	0.45	0.07	0.28	0.60
	Reserve	First	2	0.35	0.06	0.31	0.39	0.19	0.01	0.18	0.19
		Second	2	0.62	0.06	0.58	0.65	0.36	0.00	0.36	0.37
	Weak	First	1	0.27	.	0.27	0.27	0.03	.	0.03	0.03
		Second	1	0.50	.	0.50	0.50	0.10	.	0.10	0.10

*Total Number of Items

Table 3. 3 CTT Descriptive Summary of Field-Test Item Pool, Science Spring 2019

Grade/ Course	Flag	Trial	N*	P–value				Point Biserial			
				Mean	SD	Min	Max	Mean	SD	Min	Max
5	Keep	First	59	0.53	0.13	0.32	0.83	0.43	0.10	0.19	0.62
		Second	59	0.78	0.09	0.53	0.93	0.40	0.09	0.18	0.55
	Reserve	First	1	0.39	.	0.39	0.39	0.16	.	0.16	0.16
		Second	1	0.68	.	0.68	0.68	0.30	.	0.30	0.30
8	Keep	First	58	0.56	0.12	0.27	0.80	0.50	0.11	0.22	0.68
		Second	58	0.79	0.09	0.50	0.94	0.50	0.09	0.31	0.68
	Reserve	First	1	0.37	.	0.37	0.37	0.19	.	0.19	0.19
		Second	1	0.81	.	0.81	0.81	0.35	.	0.35	0.35
	Weak	First	1	0.35	.	0.35	0.35	0.01	.	0.01	0.01
		Second	1	0.74	.	0.74	0.74	0.19	.	0.19	0.19
Biology	Keep	First	58	0.51	0.13	0.26	0.83	0.41	0.10	0.20	0.61
		Second	58	0.78	0.07	0.62	0.94	0.42	0.08	0.18	0.61
	Weak	First	2	0.36	0.03	0.34	0.38	0.07	0.03	0.05	0.09
		Second	2	0.66	0.06	0.62	0.70	0.18	0.12	0.10	0.27

*Total Number of Items

3.3.1 Field-Test Sample Characteristics

Table 3.4 shows the demographic characteristics of the alternate population who participated in the standalone field test including breakdown by economically disadvantaged students (EDS) and ELs status.

Table 3.5 and *Table 3.6* show the percentage of students who participated in the standalone field test by primary eligibility area. The students whose primary eligibility area is Autism had the highest percentage of participation followed by Intellectual Disability-Moderate, Intellectual Disability-Mild, and Multiple Disabilities. The discrepancy in total student count between *Table 3.4* and *Table 3.5* or *Table 3.6* for some grades is primarily because some students' disability categories were not specifically labeled in the data set.

Table 3. 4 Demographic Characteristics of NCEXTEND1 FT Population, 2018–19

Grade/ Course	Gender (%)		Ethnicity (%)				EDS (%)	ELs (%)	Total
	Female	Male	Black	Hispanic	Other	White			
Reading									
3	31.0	69.0	33.4	19.1	10.2	37.2	61.3	14.0	1,182
4	32.1	67.9	35.6	19.5	9.9	35.0	60.1	13.7	1,224
5	30.4	69.6	36.5	17.5	9.6	36.5	60.3	12.1	639
6	33.0	67.1	35.1	16.0	10.2	38.8	60.9	10.0	1,202
7	33.2	66.9	34.5	15.3	6.9	43.3	64.8	10.0	1,246
8	32.7	67.3	36.0	14.4	6.2	43.5	59.6	10.4	550
10	31.5	68.5	34.6	12.3	8.5	44.6	58.9	6.8	457
Science									
5	33.5	66.5	35.6	17.6	8.2	38.6	59.1	10.7	624
8	34.9	65.1	34.4	16.2	7.6	41.8	61.4	11.3	593
Biology	34.3	65.7	33.0	12.9	8.7	45.5	55.1	12.1	519

Table 3. 5 NCEXTEND1 Reading FT Student Population by Disability Sub-Group, 2018–19

Disability Categories	Grades						
	3	4	5	6	7	8	10
Autism (AU)	42.83	41.00	37.50	37.08	35.69	31.45	32.39
Deaf-Blindness (DB)	0.17	0.16		0.08	0.08		0.44
Developmental Delay (DD)	0.51	0.16	0.16				
Deaf (DF)		0.08		0.17	0.16		
Serious Emotional Disability (ED)	0.17	0.58	0.47	0.33	0.32	0.36	0.22
Hearing Impairment (HI)	0.17	0.08		0.25	0.24		0.66
Intellectual Disability - Mild (IDMI)	13.99	14.54	16.46	15.17	16.96	18.00	16.63
Intellectual Disability - Moderate (IDMO)	20.70	22.68	23.10	24.58	25.32	26.73	27.35
Intellectual Disability - Severe (IDSE)	3.14	4.44	3.80	4.33	3.86	4.55	6.35
Specific Learning Disability (LD)	0.93	0.41	0.79	0.25	0.24	0.55	0.88
Multiple Disabilities (MU)	11.96	10.35	9.34	11.25	11.66	12.91	10.28
Other Health Impairment (OH)	3.99	3.94	6.49	4.33	3.46	4.55	2.84
Orthopedic Impairment (OI)	0.42	0.33	0.47	0.33	0.40		1.09
Speech or Language Impairment (SI)				0.17	0.08		
Traumatic Brain Injury (TB)	0.85	1.15	1.42	1.42	1.21	0.91	0.66
Visual Impairment (VI)	0.17	0.08		0.25	0.32		0.22
Total Number of Students	1,179	1,217	632	1,200	1,244	550	457

Note: Some students are not coded for a specific disability

Table 3. 6 NCEXTEND1 Science FT Student Population by Disability Sub-Group, 2018–19

Disability Categories	Grades		
	5	8	Biology
Autism (AU)	36.5	34.3	34.36
Deaf-Blindness (DB)		0.17	
Serious Emotional Disability (ED)	0.48	0.17	0.39
Hearing Impairment (HI)	0.48		0.19
Intellectual Disability - Mild (IDMI)	15.76	14.26	12.74
Intellectual Disability - Moderate (IDMO)	26.69	27.5	28.19
Intellectual Disability - Severe (IDSE)	4.18	6.11	4.44
Specific Learning Disability (LD)	0.48	0.68	0.19
Multiple Disabilities (MU)	10.13	10.53	14.09
Other Health Impairment (OH)	4.34	3.9	3.67
Orthopedic Impairment (OI)	0.16	0.17	0.58
Speech or Language Impairment (SI)	0.16		
Traumatic Brain Injury (TB)	0.64	2.21	0.97
Visual Impairment (VI)			0.19
Total Number of Students	622	589	518

Note: Some students are not coded for a specific disability

3.3.2 Raw Scores and Timing Data

Table 3.7 shows descriptive statistics of raw scores and time (minutes) taken by the students during the standalone field-test. The Reading results indicate students on average scored 22.5 to 23.4 points out of 36 points (18 items) across grades. Their average testing time ranged from 23.1 to 29.6 minutes indicating students on average took about 1.28 to 1.64 minutes per item to complete the test. Results further indicated that approximately 95% of the students completed the test in around 1-hour.

For science, students on average scored 25.7 to 27.0 points in a 40-point (20 items) test and took on average 15.2 to 19.1 minutes indicating less than 1-minute per item. The results further indicated that about 95% of the student completed the tests in 46.4 minutes or less.

There are many variables to consider when collecting the time data including testing in multiple days, breaks, if the testing cards were used, and how the teacher and students interact with the system.

Table 3. 7 Raw Score Descriptive Statistics and Testing Time, 2018–19 FT

Grade/ Course	N	No. of Items	Raw Score					Time (Minutes)					
			Mean	SD	50th	Min	Max	Mean	SD	50th	Min	95th	99th
Reading													
3	1119	18	23.3	7.7	24	0	36	27.5	92.3	15.9	3.2	42.7	140.3
4	1176	18	23.3	7.1	23	0	36	23.1	43.7	16.3	2.8	53.6	132.2
5	623	18	22.5	6.8	22	0	36	26.1	68.6	16.9	1.5	44.4	135.7
6	1164	18	23.2	7.5	23	0	36	26.5	70.4	17.8	2.4	47.2	131.6
7	1204	18	23.4	6.9	23	0	36	27.5	47.2	19.4	3.3	54.9	143.9
8	542	18	22.5	6.8	22	0	36	29.6	77	19.6	2.3	71.9	136.0
10	448	18	22.5	6.7	22	0	36	28.5	31.9	22.6	2.8	50.5	143.0
Science													
5	596	20	26.0	6.9	26	0	40	18.3	24.8	12.2	16.3	40.7	135.7
8	579	20	27.0	8.2	28	0	40	19.1	23.8	13.1	18.0	46.4	128.0
Biology	506	20	25.7	6.8	26	0	40	15.2	17.8	11.9	15.7	26.6	118.8

Note: Time data for some students was not clear and was removed from the calculation.

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. A fairness review panel was convened to examine all items. As a developer of the NCEXTEND1 assessments, it is the responsibility of the 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, the 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 working with 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 a bias item, sensitive to bias, 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 TMS, Psychometrician, and Lead Content Specialist at NCSU-TOPS. The less-than-optimal items were allowed to be included 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 form assembly processes used to create NCEXTEND1 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 24–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

Edition 4 NCEXTEND1 alternate assessments were developed manually 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. Moreover, only three achievement levels were used in *Edition 4* as opposed to five levels in *Edition 3*.

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. Notice that the average p-value for the first trial ranged from 0.49 to 0.52 and second trial ranged from 0.75 to 0.78 across grades. The point biserial ranged from 0.44 to 0.49 for first trial and 0.43 to 0.49 for second trial. The NCEXTEND1 tests are teacher administered by scaffolding of item options.

Table 4.1 CTT Summary Statistics of New Form Based on 2018–19 FT

Grade/ Course	No. of Items	First Trial		Second Trial	
		P-Value	Point Biserial	P-Value	Point Biserial
3	24	0.49	0.46	0.75	0.49
4	24	0.49	0.47	0.75	0.47
5	24	0.50	0.46	0.76	0.44
6	24	0.51	0.49	0.76	0.47
7	24	0.52	0.45	0.78	0.43
8	24	0.50	0.44	0.76	0.46
10	24	0.49	0.46	0.76	0.43
Science					
5	25	0.51	0.47	0.78	0.40
8	25	0.52	0.48	0.77	0.49
Biology	25	0.51	0.43	0.78	0.43

4.2 Form Review

The NCEXTEND1 operational form review is a 24–step process (*Appendix 2–D, p. 25–28*). After the initial assembly and statistical characteristics (Step 1) are satisfied, 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 process. If there is a replacement item from the item bank that maintains the blueprint and statistical properties of the form, then a swap is made, and the form sent back through the review process.

4.2.1 Content Reviews

In the NCEXTEND1 operational form review processes, the content review steps are Steps 3, 5, 6, 8, 9, 14, 16, 17, and 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 documents his or her opinion as to how well the tests met the five criteria listed above. These comments were further reviewed by the NCSU-TOPS and the NCDPI's test measurement 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 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 Reading 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 for 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 Test Development System (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 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.

The approved forms are exported to Steps 22-24 for final freeze and approval. At the steps, 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 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 socioeconomic status, culture, and language considerations.

The 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 reviewers is an EC educator, and the other is a general education. The EC educator 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

The reviewers bring different expertise to the group when determining whether forms are accessible to students participating on the NCEXTEND1 alternate assessments. The EC reviewer analyzes items for accessibility concerns as they relate to students with significant cognitive disabilities. The English as a Second Language (ESL) reviewer looks at the items as they relate to students who are dually identified as a student with a disability and as an English Learner. Lastly, the VI reviewer analyzes the items for accessibility for students with little or no vision in terms of graphics and the use of braille. 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 socioeconomic reasons. Review of the 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, and
- Look for idioms that may provide an accessibility issue.

Detail bias and sensitivity review processes are documented in *Appendix 3–A*. In 2018–19 administration, the Fairness Review panel for both general and NCEXTEND1 was made up of 11 participants representing teachers and educators. These members were selectively recruited based on their expert knowledge of 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 Education Moodle courses.

Table 4.2 Demographic Information for Fairness Review Panels, 2018–19

Category	Subcategory	N	%
Gender	Female	5	45%
	Male	6	55%
Ethnicity	African American	4	36%
	Asian	1	9%
	White	3	27%
	Hispanic	1	9%
	Native American	1	9%
	Other	1	9%
Highest Degrees Earned	BA/BS	4	36%
	MA/MS	5	45%
	Other	1	9%
	Ph. D	1	9%
Year of Experience	>20	6	55%
	10–20	3	27%
	1–10	2	18%

The items selected in the operational reading and science NCEXTEND1 forms are cleared from the bias and sensitivity reviews.

4.3 Summary of Final Operational Forms and Field-Test Design

This section details test structures and statistical properties of new *Edition 4* NCEXTEND1 Reading and Science alternate assessments that were built in 2019 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 Reading and Science Operational Form Structures

The NCEXTEND1 Reading Alternate Assessments are teacher-facilitated online assessments that are administered individually to each student. The questions are presented online in two (2) sets. There is no formalized break between the sets. The item format is 3-option selected response presented online in two (2) sets. Rationales for using 3-option selected response items are driven by practical and policy considerations. Scaffolding methods are used throughout instruction and classroom tasks to engage students with significant cognitive disability. 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.

Reading at Grades 3–8 and English II at Grade 10 Operational Form Structures: The NCEXTEND1 operational assessment structures for Reading at grades 3–8 and 10 included 30 performance-based, 3-option multiple-choice items: six selections—five selections with 24 operational items and one selection with six field-test items not to be included in the score but used for purposes of developing future test forms (*Table 4.3*). All selections from Set 1 are read by the teacher. For Set 2, one operational and one embedded selection are read by the teacher, and one operational selection is read by the student. The NCEXTEND1 alternate assessments are administered individually to each student.

Table 4.3 NCEXTEND1 Reading Operational Test Structure

Grade/Course	Operational		Embedded Field Test Items		Total	
	Operational		Field-Test		Total	
	Selections	Items	Selections	Items	Selections	Items
Grades 3–8 and 10	Set 1: 3	14	—	—	3	14
	Set 2: 2	10	Set 2: 1	6	3	16
	Total	24		6	6	30

Science at Grades 5 and 8 and Biology at grade 10 Operational Form Structures: *Table 4.4* shows the operational test structures for the NCEXTEND1 Science alternate assessments. Each operational form comprised of a total of 30 items— 25 operational and five field-test. Set 1 consis of 12 operational and two field test items and Set 2 consist of 13 operational and three field test items.

Table 4.4 NCEXTEND1 Science Operational Test Structure

Grades/Course	Set	Operational	Field-Test	Total
Grades 5, 8, and Biology	Set 1	12	2	14
	Set 2	13	3	16
	Total	25	5	30

4.3.2. Field-Test Plan

Each operational form of NCEXTEND1 Reading test has six (6) embedded field test slots in Set 2 of the test (*Table 4.3*). Similarly, NCEXTEND1 science test has five (5) embedded field test slots - two in Set 1 and three in Set 2 (*Table 4.4*). At the conclusion of every testing cycle, the NCDPI Psychometric team shall evaluate item statistics of the base form for possible item drift. If the NCDPI decided to develop a new NCEXTEND1 form or replace some of the items from the current form for succeeding administration, up to two flavors shall be created to embed new items.

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 Reading and Science 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 NCEXTEND1 alternate assessment. The guides for assessors and coordinators are available to ensure standardized administration of all NCEXTEND1 alternate assessment given across the state. They are briefly described below with website links for more detailed descriptions.

NCEXTEND1 Alternate Assessment Guide: The Reading and Science NCEXTEND1 alternate assessments measure corresponding ECSs. The assessments were redesigned and adopted by the NCSBE in 2017–18 administration. The assessments are administered online to all eligible students at the appropriate grade levels. The alternate assessment guide contains the administrative procedures applied to the NCEXTEND1 test administration and posted online.

Testing Students with Disabilities Handbook*:* This document contains policy guidelines and procedures for testing students with disabilities, including students with significant cognitive disabilities, in the North Carolina Annual Testing Program. Students who are classified as having a significant cognitive disability are those who receive special education and related services under the *Individuals with Disabilities Education Improvement Act of 2004 (IDEA)* School personnel must ensure the policy guidelines and procedures outlined in this publication are implemented appropriately (*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 from *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 ECSs in Reading and North Carolina Extended Essential Standards for Science.
- The student must be enrolled in grades 3–8, 10, or 11, according to PowerSchool.

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

- are being instructed in any or all the assessed general grade- or course-level content standards of the North Carolina *Standard Course of Study* in Reading and the Essential Standards in Science,
- demonstrate delays only in academic achievement,
- demonstrate delays only in selected areas of academic achievement,
- demonstrate delays owning 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.

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. 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. Schools will have the option to order test cards when administering the NCEXTEND1 tests. Test cards are available for all grades and all subjects. Schools opting to not use test cards, however, will need to order a test ticket for all NCEXTEND1 tests. Test cards and test tickets will be ordered through the Testing News Network (TNN) ordering page. The decision to use the cards should be 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 of the *NCEXTEND1 Alternate Assessment Guide*. There are three (3) administration options available for the Reading and Science NCEXTEND1 alternate assessment. 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 NCEXTEND1 alternate assessment 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 NCEXTEND1 tests for 2020–21 and beyond are computer-based with three answer recording options as stated in Section 5.4.2. Before the administration of the NCEXTEND1 alternate assessment, 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 assessment should complete 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, one grade-appropriate test card kit per student assessed (if needed), a supply of scratch paper, and sharpened pencils with erasers or other student-specific writing utensils.

5.6 Technical Specifications for NCEXTEND1 Assessment Administration

To ensure students receive valid and reliable assessment administrations, schools must meet specific technical requirements. The devices that will be used for the online 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 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:

- Verify that the assessor's NC Education username and password is working correctly. The assessor should ensure accessibility to the course tabs in NCTest Admin. Usernames and passwords should be checked two to three days before the test administration date. Passwords are never shared, and if compromised, the school test coordinator must be notified immediately.
- 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 NCEXTEND1 alternate assessment 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 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, medical condition, and/or fatigue factor(s). The NCEXTEND1 alternate assessment 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.). There is no formalized break in the sets. The test design for NCEXTEND1 alternate assessment 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 2020–21 NCEXTEND1 operational Reading and science assessments are shown in *Table 5.1.* The average time taken by the students to complete the operational test is about 37 minutes or less—about one minute per item for Reading and 19 minutes or less for the 30 items test for Science. Moreover, 95% of students completed the Reading assessments within approximately 81 minutes. For science, 95% of the students completed the 30 items test in 36 minutes or less. Notice that the standard deviation (SD) of time in minutes is large indicating high variation of time examinees' taking to complete the tests.

Table 5. 1 Recorded Test Duration (Minutes) for Reading and Science Operational Forms, 2020–21

Grade/ Course	N	Number of Items	Statistics		Percentile				
			Average	SD	25th	Median	75th	95th	99th
Reading									
3	1,078	30	30	36	21	27	36	66	195
4	1,102	30	29	29	21	27	33	57	123
5	1,167	30	32	33	21	27	36	66	165
6	1,116	30	29	39	21	30	36	60	141
7	1,104	30	31	38	24	30	36	63	144
8	1,235	30	34	36	24	33	42	72	150
10	876	30	37	44	27	36	45	81	189
Science									
5	1,136	30	19	16	12	15	21	36	66
8	1,221	30	19	18	12	18	21	36	63
Biology	880	30	17	20	12	15	18	33	87

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 assessment 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 may be provided with appropriate administrative procedures to assure that individual student needs are met while maintaining sufficient uniformity of the test administration. Approved accommodations for the NCEXTEND1 alternate assessment 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 and 10	Science Grades 5, 8, and Biology
Braille Materials (Full UEB)	Yes	Yes
Braille Writer/Braille Paper	Yes	Yes
Large Print Materials	Yes	Yes
Assistive Technology Devices	Yes	Yes
Interpreter/Transliterator Signs/Cues Test	NCEXTEND1 tests are to be read aloud to all students as specified in the NCEXTEND1 Assessment Guide	
Magnification Devices	Yes	Yes
Word-to-Word Bilingual (English/Native Language) Dictionary/Electronic Translator (ELs only)	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
Testing in a Separate Room	Yes	Yes
Adaptations to NCDPI-Provided Manipulatives	Yes	Yes

For more information on accommodations, refer to the *Testing Students with Disabilities Handbook*.

5.9 Student Participation

For a student with a significant cognitive disability to participate in a 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 ECSs. 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 ECSs. 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,

- who requires extensive and repeated individualized instruction and support to make meaningful gains, and
- who 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 2020–21 Reading and science general and the NCEXTEND1 alternate assessment are presented in *Table 5.3*. 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, Spring 2021

Grade/ Course	General		NCEXTEND1		Total N
	N	%	N	%	
Reading					
3	105,890	98.8%	1,242	1.2%	107,132
4	106,164	98.8%	1,281	1.2%	107,445
5	108,163	98.8%	1,340	1.2%	109,503
6	110,928	98.9%	1,215	1.1%	112,143
7	112,125	98.9%	1,211	1.1%	113,336
8	112,293	98.8%	1,345	1.2%	113,638
10	109,807	99.2%	917	0.8%	110,724
Science					
5	107,898	98.8%	1308	1.2%	109206
8	111,690	98.8%	1330	1.2%	113020
Biology	104,690	99.1%	919	0.9%	105609

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 a testing exception for the student from the NCDPI. 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. Request for Testing Exceptions Based on Significant Medical Emergencies and/or Conditions can be accessed from the NCDPI website (*Appendix 5–E).*

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. The 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*.

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 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–F*)

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 materials not stored in secure locked area
- Items from secure test used for instruction
- 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
- 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.

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 10 Reading, grades 5 and 8 science, and Biology NCEXTEND1 alternate assessment 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 grades 3–8 and 10 Reading NCEXTEND1 alternate assessments' 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 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 population and likely small variation of abilities in the population, use of IRT models was not appropriate for the assessments. Instead, raw scores were linearly transformed to reportable scale scores using expected 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 NCEXTEND1 alternate assessment. The use of the pre-established scoring tables allows the NCDPI to take full advantage of short administration window for the NCEXTEND1 alternate assessment 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 classical summary statistics based on the operational 2020–21 administration are shown in *Table 6.1*. The average P-value and point-biserial for the Spring 2021 population varied albeit minimally to the one obtained from the field-test in Spring 2019 (*Table 4.1*) with differences ranging only from -0.01 to 0.03.

The small difference between the field-test and operational statistics indicates the item statistics are relatively stable.

Table 6. 1 Average CTT Statistics, 2020–21 Operational Forms

Grade/ Course	N	No. of Items	First Trial		Second Trial	
			P-Value	Point Biserial	P-Value	Point Biserial
Reading						
3	~1054	24	0.51	0.49	0.75	0.52
4	~1083	24	0.49	0.45	0.74	0.47
5	~1157	24	0.52	0.45	0.78	0.45
6	~1094	24	0.54	0.49	0.77	0.49
7	~1080	24	0.53	0.47	0.78	0.48
8	~1216	24	0.50	0.44	0.76	0.47
10	~860	24	0.48	0.45	0.75	0.49
Science						
5	~1138	25	0.54	0.47	0.79	0.46
8	~1220	25	0.54	0.46	0.77	0.49
Biology	~876	25	0.54	0.47	0.78	0.52

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

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, the NCDPI used linearly transformed scale score from raw score for the Reading and science NCEXTEND1 alternate assessment reporting. The transformation included setting expected mean and standard deviation of the 2020–21 population. For the NCEXTEND1 scaling procedures the following linear transformation of raw scores were implemented:

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

Where Yx 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 2020–21 operational administration, and Mean(e) is the expected mean of the scale score. The expected theoretical means of the scale scores for Reading at grades 3–8 and 10 were set to be 430, 440, 450, 460, 470, 480, and 490 with standard deviation of 9. Similarly, expected means for the Science at grades 5 and 8, and Biology at grade 10 were set to be 450, 480, and 490 with standard deviation of 9. During the base year (2020–21) of implementation of *Edition 4* ECSs, 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 rational and procedures used for establishing cut score should be documented*" (p. 107). For the first operational administration of the *Edition 4* NCEXTEND1 2020–21 Reading assessments, NCDPI contracted with the Data Recognition Corp (DRC[1]) 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 are 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 evidence in abbreviated version of the *Edition 4* NCEXTEND1 Reading and Science 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.*

In school year 2019–20, the North Carolina tests of NCEXTEND1 Reading and Science transitioned to measuring students' command of the North Carolina ECSs. Test designs for NCEXTEND1 were also modified to improve measurement precision. The rigor associated with the NCEXTEND1 assessment was also enhanced to align with expectations outlined in the ECSs. These changes to the tests were put in place for planned testing in spring 2020; however, spring 2020 testing was canceled due to the COVID-19 pandemic. These changes were implemented in school year 2020–21 when testing resumed. In response to the changes to the standards and test structures, the NCDPI sponsored a standard setting for the NCEXTEND1 Reading and Science.

[1]Copyright © 2019 Data Recognition Corp.

The standard setting took place in three components over a five-day period. Participants used the Angoff Yes/No process to recommend cut scores. This procedure has been used to establish performance standards for educational assessments in North Carolina and around the world.

7.1 Standard Setting Activities

The standard setting for Reading and Science NCEXTEND1 Alternate Assessments was originally planned for the summer of 2020 based on 2019–20 operational data. Due to the COVID-19 pandemic, all Spring 2020 test administrations were canceled. The Spring 2021 data were post-pandemic with ununiform instructional practices. Therefore, item statistics from Spring 2019 standalone field-test were used for the standard setting. The results from the standard setting will be monitored in future administrations.

Reading: Six educators involved in NCEXTEND1 Reading ALD writing at grade 8 and 10 on July 12, 2021. Twenty-four educators involved in the standard setting for NCEXTEND1 Reading at grades 3–8 and 10 on July 13-16, 2021. Of the 24 educators, 11 worked on grades 3–6 and 13 on grades 6–8 and 10. The educators focused on grade 6 before diverging into two groups to focus on grades 3–5 and on grades 7–8 and 10.

Science: Twenty-four educators involved in the standard setting for NCEXTEND1 Science at grades 5 and 8 and Biology at grade 10 on July 12-13, 2021. Of the 24 educators, 12 participants worked on grades 5 and 8, and 12 participants worked on grade 8 and 10. Participants focused on grade 8 before diverging on grade 5 and Biology.

The goal of the workshop was to identify cut scores that divide students into three achievement levels: Not Proficient, Level 3, and Level 4.

7.1.1 Participants' Characteristics

The gender and ethnic composition of the participants indicate there were majority female and White (*Table 7.1*). The self-reported years of experience shows the majority of participants served 11 years or more in the profession (*Table 7.2*). Moreover, the majority were general education teachers followed by special education teachers (*Table 7.3*). Some participants were curriculum staff and District-Level administrators.

Table 7.1 Self-Reported Gender and Ethnicity

Content Area	N	Gender		Ethnicity			
		Female	Male	White	Black	Asian	No Response
Reading	24	23	1	20	3	1	0
Science	24	23	1	16	6	1	1

Table 7 2 Self-Reported Years of Experience

Content Area	N	<5	5–10	11–15	16–20	21–25	>25
Reading	24	1	3	6	7	2	5
Science	24	2	7	2	5	5	3

Table 7.3 Self-Reported Current Position

Content Area	N	General Education	Special Education	Curriculum Staff	District-Level Administrator
Reading	24	10	6	6	2
Science	24	12	9	2	1

7.1.2 Opening Session and Introductions

On each workshop day, all participants began the workshop with a single opening session led by the NCDPI. During these sessions, the NCDPI's Director of Accountability Services welcomed the participants to the workshop, described the purpose of the workshop, and subsequently described the recent changes to the North Carolina standards and tests, and how valuable the participating educators' recommendations would be in identifying new cut scores for the tests.

Following committee introductions, each grade level panels spent the remainder of the day discussing 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 panelists 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 reading and science 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 the ALD development process by drafting the initial ALDs, completed rounds of webinars, and worked on revisions with 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 7.4*) 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. Level 4 students are also considered on track for competitive employment and post-secondary education.

Table 7.4 Policy ALDs for NCEXTEND1 Alternate Reading Assessment

Not Proficient	Level 3	Level 4
Students who are not proficient demonstrate **inconsistent understanding** of the North Carolina Extended Content Standards and will need significant support at the next grade/course.	Students at Level 3 demonstrate **sufficient understanding** of the North Carolina Extended 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 post-secondary 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). The NCDPI provided policy ALDs for the NCEXTEND1 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*.

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 the NCDPI and its advisors would be reviewing their cut score recommendations and that adjustments may be made to the cut scores by the NCDPI for policy-related reasons.

After the revision, the final achievement level Cuts (*Table 7.6*) were presented to the NCSBE on August 5, 2021 for consideration. After deliberation, the NCSBE approved the cut scores on August 5, 2021. The corresponding raw score ranges for the NCEXTEND1 achievement level cuts are shown in *Table 7.7*.

Table 7.5 Final NCEXTEND1 Recommended Cuts and Proficiency Distributions

Grade/ Course	Recommended Cuts		Percent (%) of Students in Each Achievement Level Based on Recommended Cut Scores		
	Level 3	Level 4	Not Proficient	Level 3	Level 4
Reading					
3	433	444	58.6%	30.1%	11.3%
4	443	454	64.0%	28.8%	7.2%
5	452	463	55.9%	35.5%	8.6%
6	463	474	57.7%	31.5%	10.8%
7	473	483	57.8%	31.8%	10.5%
8	483	494	60.1%	33.0%	6.9%
10	494	505	62.4%	32.3%	5.2%
Science					
5	449	460	43.1%	36.4%	20.6%
8	478	490	40.0%	42.9%	17.1%
Biology	489	502	41.8%	45.5%	12.7%

Table 7.6 NCEXTEND1 Raw Score Ranges Across Achievement Levels

Grade/ Course	Not Proficient		Level 3		Level 4	
	Min	Max	Min	Max	Min	Max
Reading						
3	0	32	33	43	44	48
4	0	32	33	43	44	48
5	0	32	33	43	44	48
6	0	33	34	44	45	48
7	0	33	34	43	44	48
8	0	32	33	43	44	48
10	0	32	33	43	44	48
Science						
5	0	31	32	42	43	50
8	0	30	31	42	43	50
Biology	0	31	32	44	45	50

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* for grades 3–8 and 10and grades 5, 8, and 10 Science. Generally, 89% 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 perceived that their opinions were valued. They agreed that the final recommended cut scores were defensible.

Table 7.7 NCEXTEND1 Standard Setting Workshop Evaluation Results (n/%)

Statement	Strongly Disagree	Disagree	Agree	Strongly Agree	Agree + Strongly Agree
The achievement standards provide a reasonable profile of performance at each level.	0%	1(2%)	26(54%)	21(44%)	47(98%)
My opinions were valued by my group.	0%	0%	14(29%)	34(71%)	48(100%)
The descriptions of the threshold students were useful during the process	0%	2(4%)	15(31%)	31(65%)	46(96%)
The facilitator provided clear instructions.	1(2%)	4(8%)	16(33%)	27(56%)	43(89%)
I believe this process will yield defensible cut scores	0%	2(4%)	19(40%)	27(56%)	46(96%)

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 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

The instructional and assessment contexts surrounding the 2020–21 school year in North Carolina varied in terms of instructional practices, for example, in-person, remote, and mixed instructional format; waiver for testing in 2019–20 and accountability reporting in 2020–21; and varying participation rates across schools and districts. Therefore, the NCDPI urges caution for interpreting summary results presented in this chapter for comparison. Furthermore, one should be cautious for referencing results from 2020–21 in future administrations as the contexts are likely to vary.

With the above context, this chapter presents test level summary results for the alternate Reading assessments based on reported scale scores and achievement levels from 2020–21 NCEXTEND1 Reading 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 assessment. *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 grades 3–8 and 10 Reading NCEXTEND1 results from 2020–21 operational administration are summarized in *Figure 8.1* through *Figure 8.10*. The results are based on all eligible students enrolled at the grade level NCEXTEND1 alternate assessment.

The expected scaling parameters for the NCEXTEND1 raw to scale linear transformation across grades 3–8 and 10 Reading were set to 430, 440, 450, 460, 470, 480, and 490 respectively with a standard deviation of 9. Similarly, expected means for the science grade 5 and 8, and Biology were set to 450, 480, and 490 with standard deviation of 9. Note, however, that *Edition 4* NCEXTEND1 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 for 2020–21 population across grades and courses have similar distributional properties corresponding to the set expected mean and standard deviation.

Figure 8. 1 Grade 3 NCEXTEND1 Reading Scale Score Distribution, Spring 2021

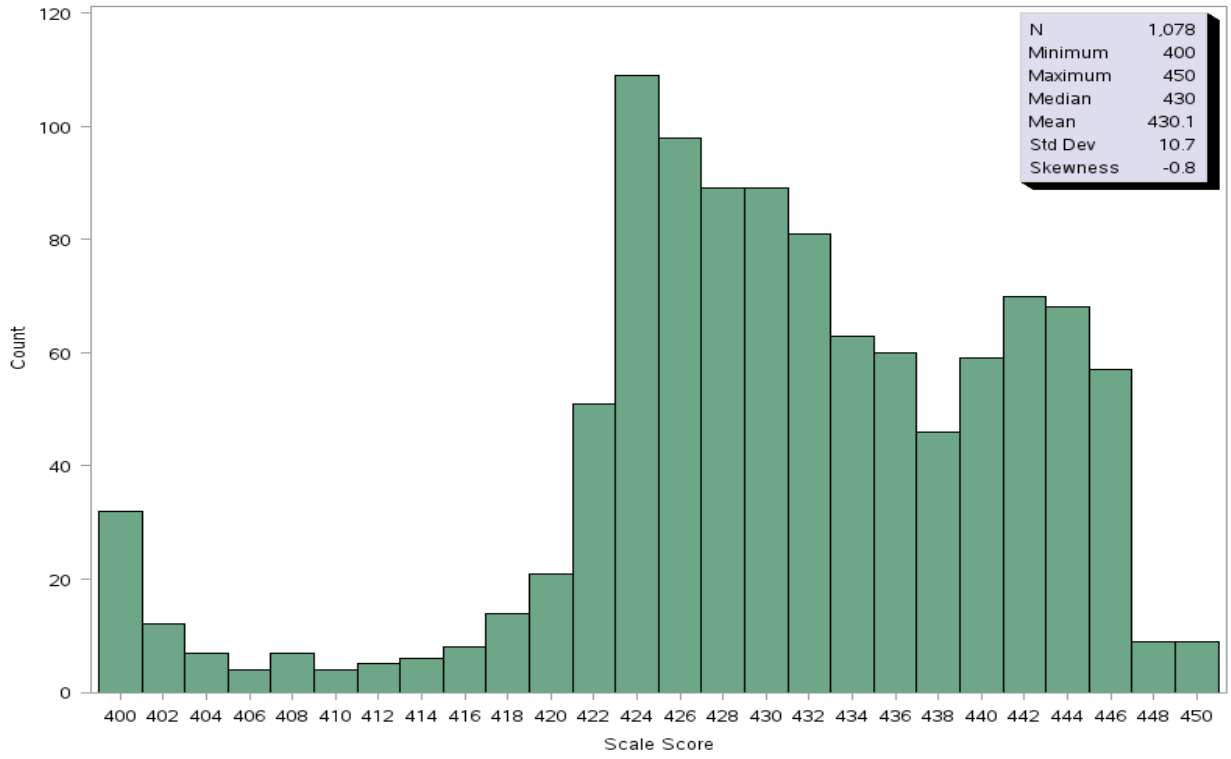


Figure 8. 2 Grade 4 NCEXTEND1 Reading Scale Score Distribution, Spring 2021

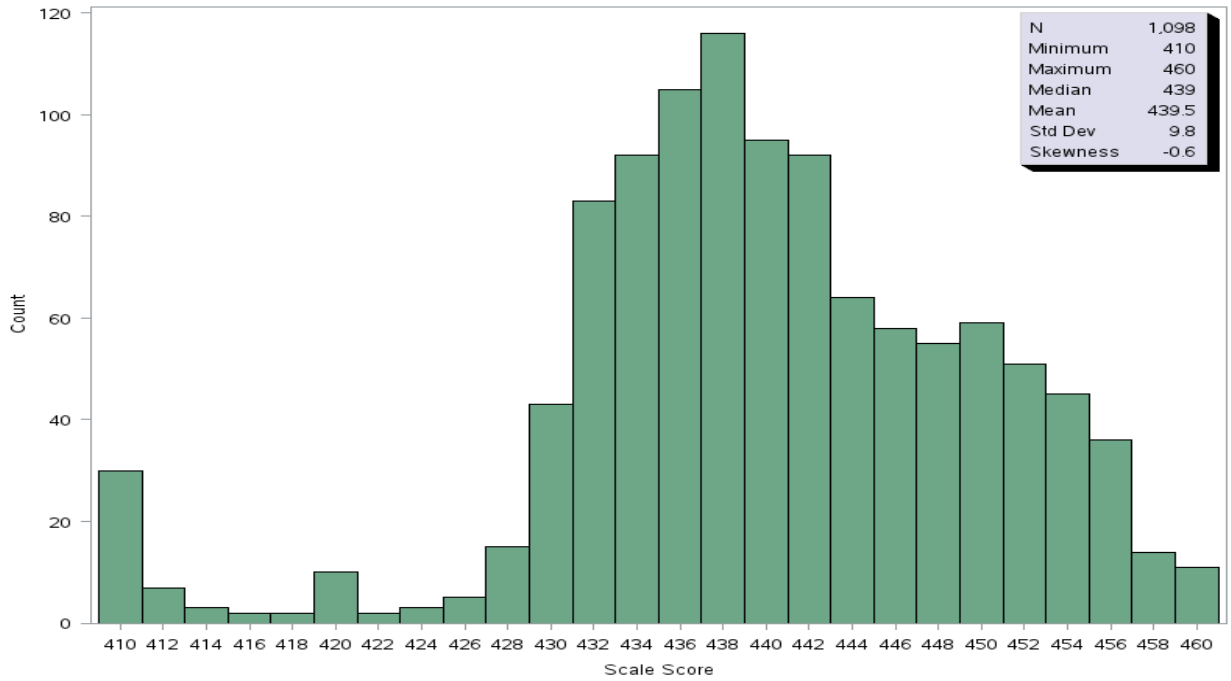


Figure 8. 3 Grade 5 NCEXTEND1 Reading Scale Score Distribution, Spring 2021

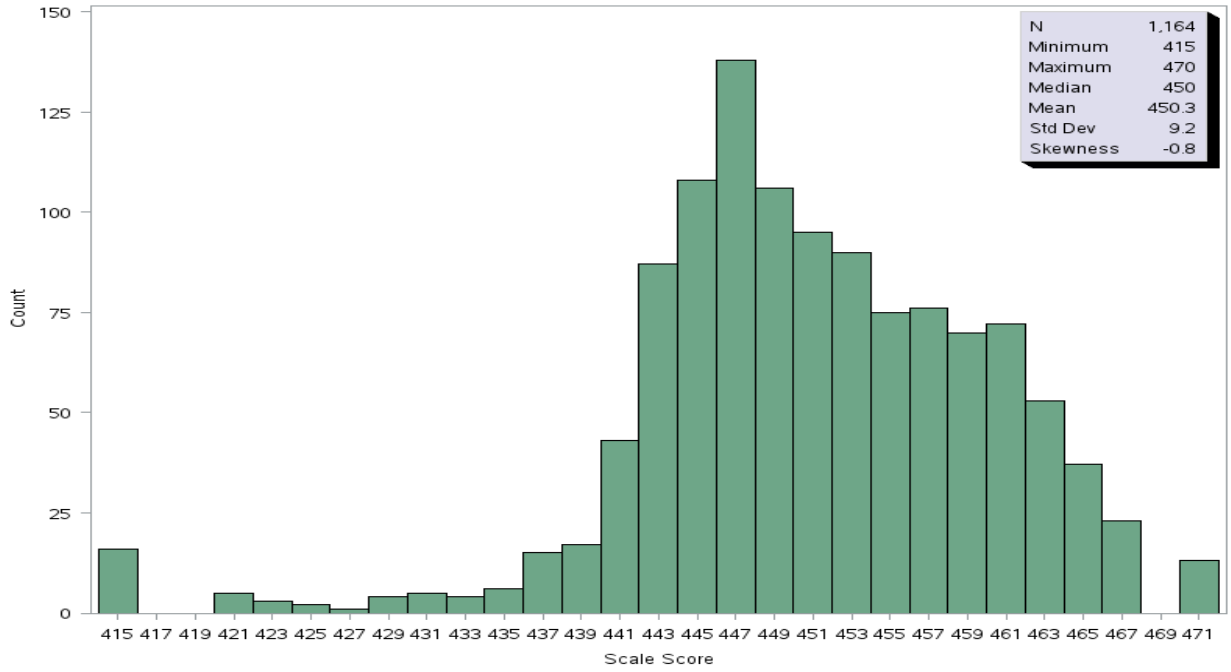


Figure 8. 4 Grade 6 NCEXTEND1 Reading Scale Score Distribution, Spring 2021

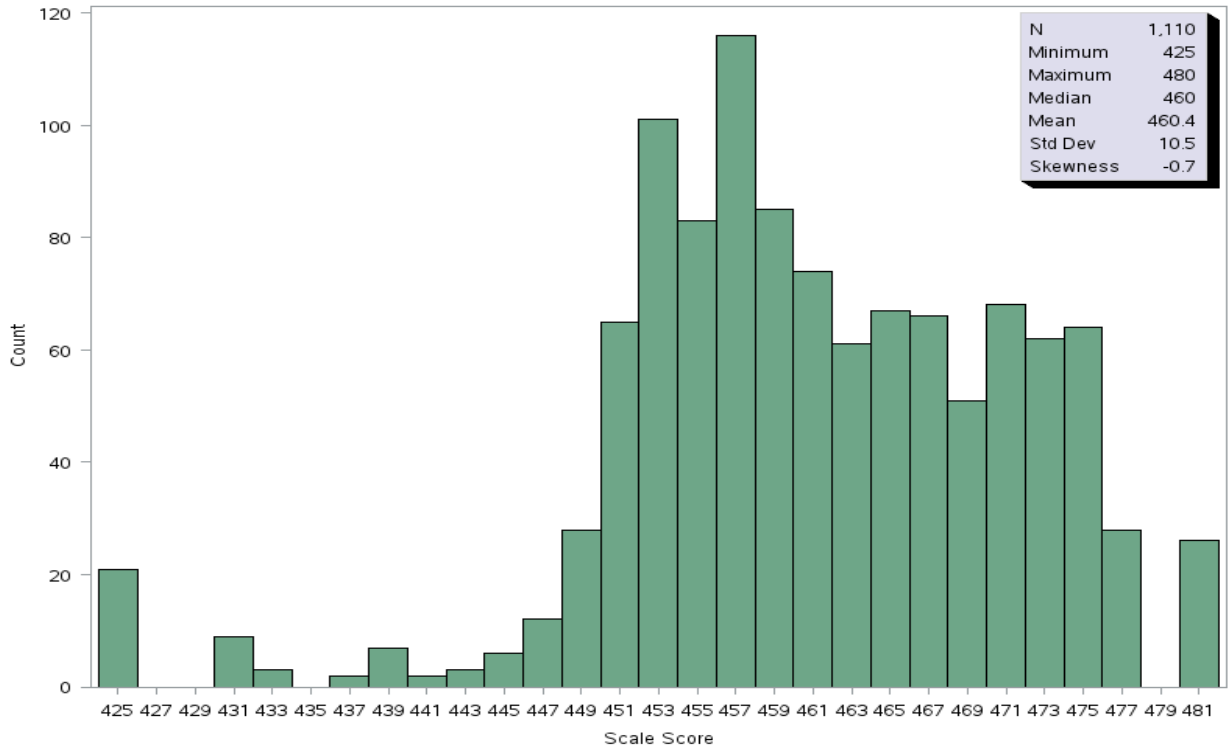


Figure 8. 5 Grade 7 NCEXTEND1 Reading Scale Score Distribution, Spring 2021

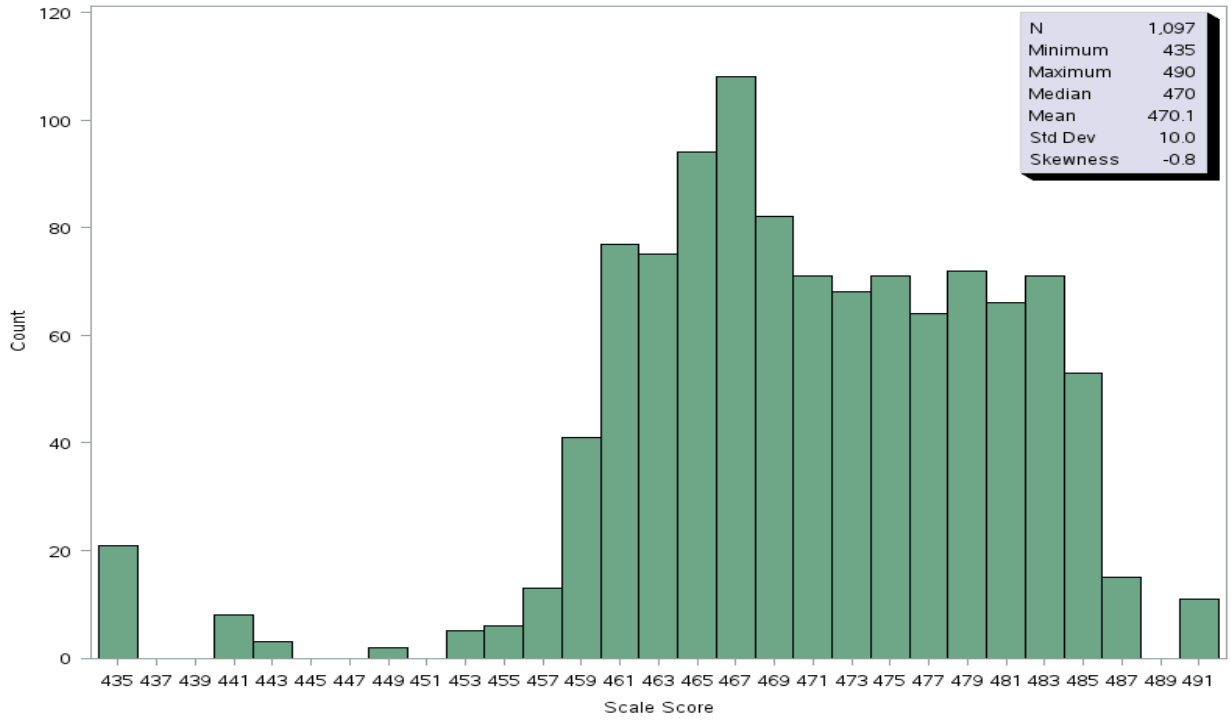


Figure 8. 6 Grade 8 NCEXTEND1 Reading Scale Score Distribution, Spring 2021

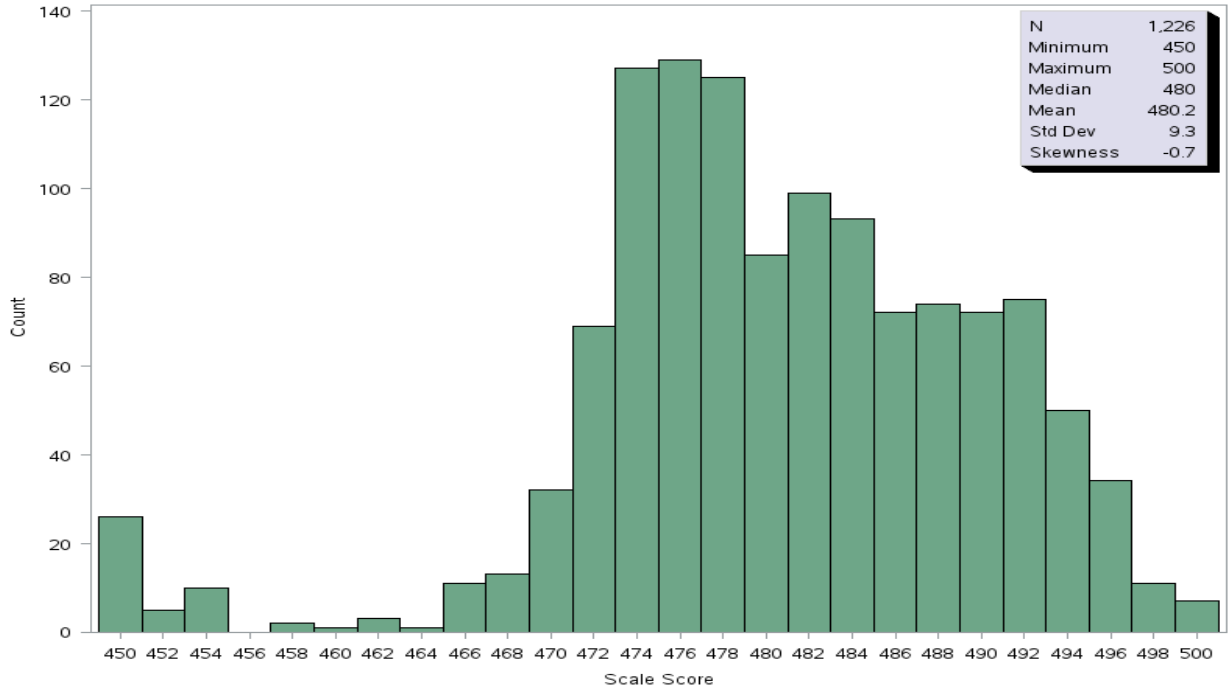


Figure 8. 7 NCEXTEND1 English II Scale Score Distribution, Spring 2021

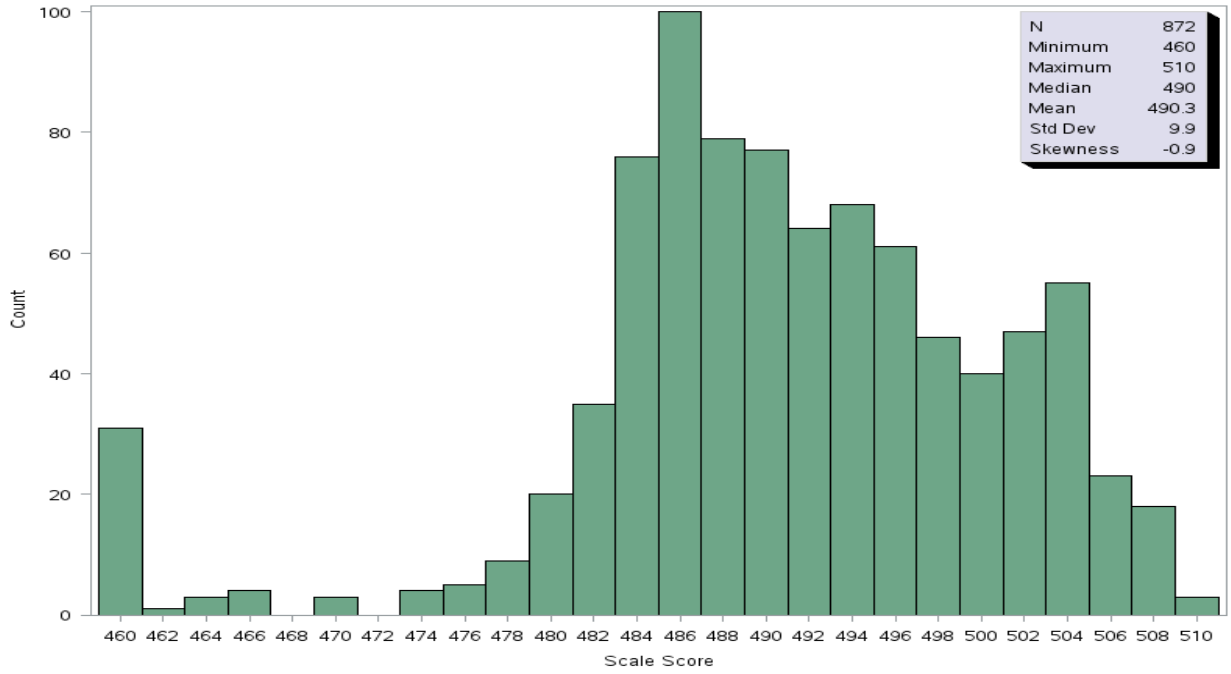


Figure 8. 8 Grade 5 NCEXTEND1 Science Scale Score Distribution, Spring 2021

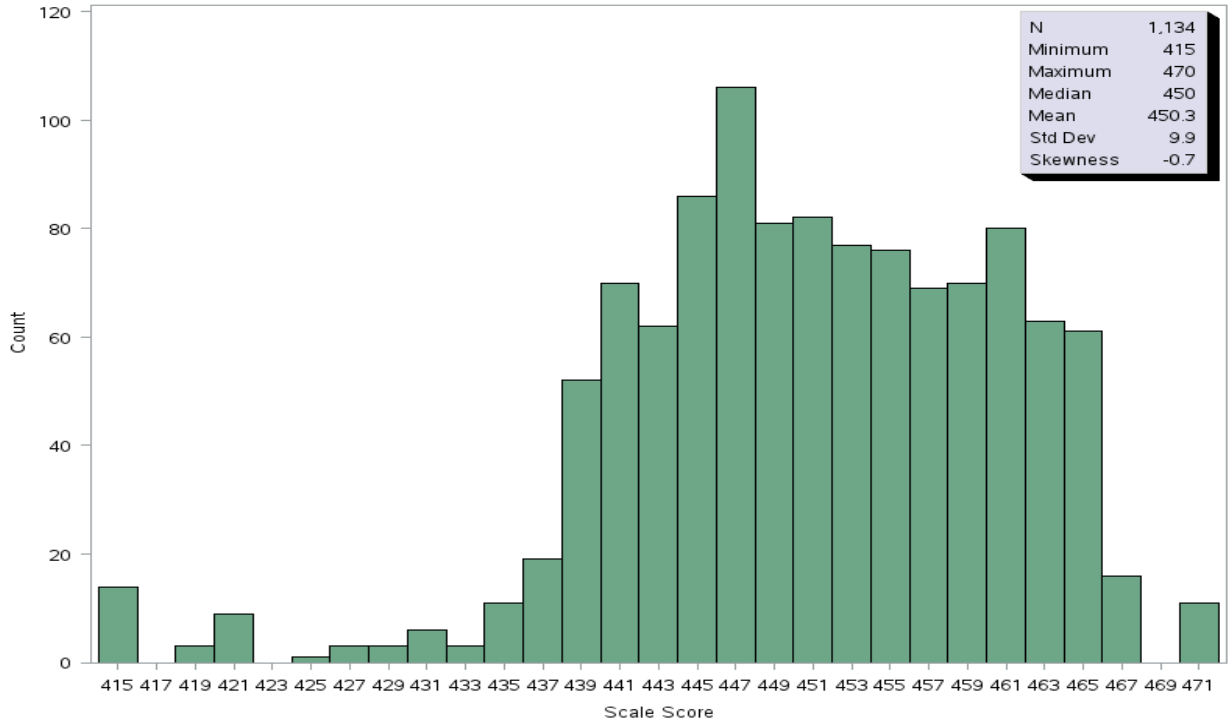


Figure 8. 9 Grade 8 NCEXTEND1 Science Scale Score Distribution, Spring 2021

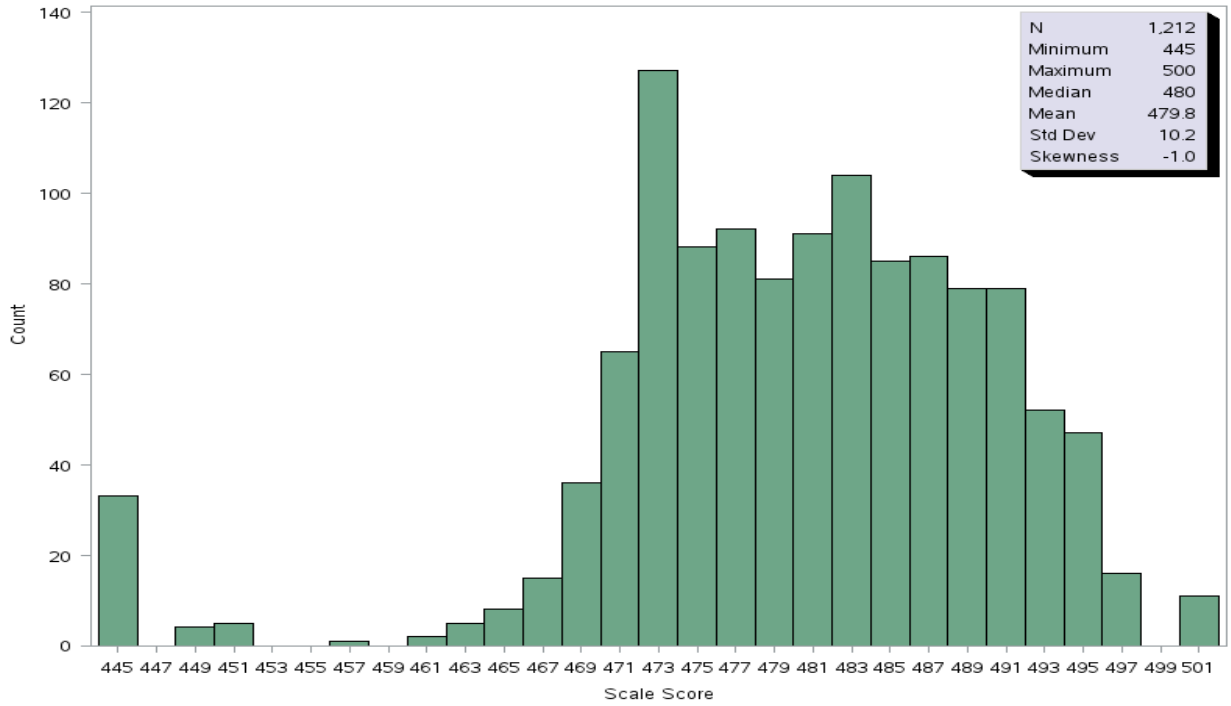
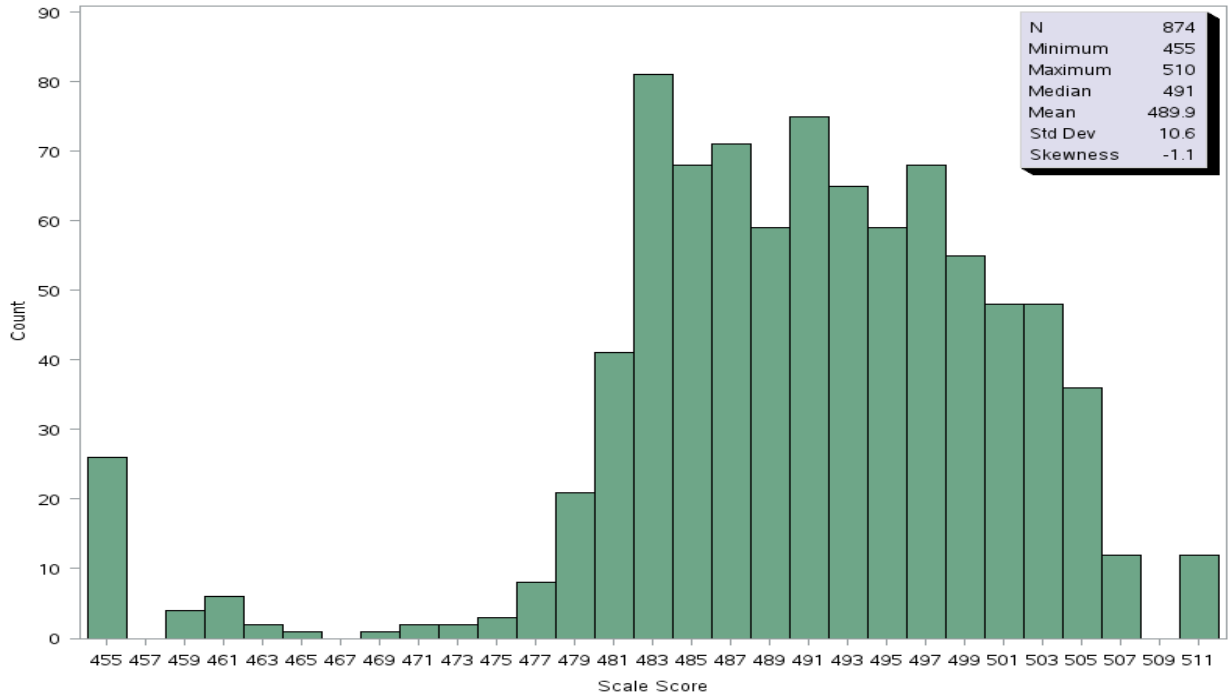


Figure 8. 10 NCEXTEND1 Biology Scale Score Distribution, Spring 2021



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 Reading at grades 3–5 are shown in *Table 8.1*, grades 6–8 and 10 are shown in *Table 8.2,* and Science at grades 5 and 8 and Biology are shown in *Table 8.3*. The major accommodation subgroups containing approximately 10% or more students are summarized separately. Those less than 10% are combined and label as "Other".

Table 8. 1 Reading at Grades 3–5 NCEXTEND1 Scale Score by Accommodation Subgroups, Spring 2021

Grade	Students with Disability Category	N	Statistics		Range		Percentile		
			Average	SD	Min	Max	25th	Median	75th
3	Autism	501	429.0	10.7	400	450	424	429	436
	Intellectual Disability - Mild	128	436.9	7.3	420	450	431	439	443
	Intellectual Disability - Moderate	181	431.7	9.0	400	450	426	431	439
	Multiple Disability	116	425.3	11.5	400	447	422	427	432
	Other	152	429.9	11.6	400	450	424	430	440
	All	1,078	430.1	10.7	400	450	424	430	439
4	Autism	471	438.7	9.9	410	460	433	439	445
	Intellectual Disability - Mild	143	445.8	8.1	420	460	439	447	452
	Intellectual Disability - Moderate	242	439.8	7.3	410	460	435	438	444
	Multiple Disability	116	435.1	11.7	410	457	430	436	443
	Other	126	439.0	10.6	410	460	434	439	445
	All	1,098	439.5	9.8	410	460	434	439	446
5	Autism	475	450.4	9.5	415	470	445	450	457
	Intellectual Disability - Mild	174	455.6	6.4	441	470	451	456	460
	Intellectual Disability - Moderate	256	449.1	7.0	415	466	444	449	453
	Multiple Disability	128	445.6	9.8	415	466	441	447	451
	Other	131	449.9	11.2	415	470	444	450	459
	All	1,164	450.3	9.2	415	470	445	450	457

Table 8. 2 Reading at Grades 6–8 and 10 NCEXTEND1 Scale Score by Accommodation Subgroups, Spring 2021

Grade/ Course	Students with Disability Category	N	Statistics		Range		Percentile		
			Average	SD	Min	Max	25th	Median	75th
6	Autism	428	459.1	10.8	425	480	453	458	466
	Intellectual Disability - Mild	162	465.8	8.4	425	480	460	467	472
	Intellectual Disability - Moderate	267	460.6	8.1	431	480	455	459	467
	Multiple Disability	106	455.5	12.8	425	480	452	456	462
	Other	147	461.3	11.0	425	480	454	460	472
	All	1,110	460.4	10.5	425	480	454	460	468
7	Autism	380	468.1	9.8	435	490	463	468	475
	Intellectual Disability - Mild	182	477.1	6.7	459	490	472	478	482
	Intellectual Disability - Moderate	289	469.8	8.6	435	490	463	470	476
	Multiple Disability	96	466.3	8.6	435	486	463	466	470
	Other	150	469.8	13.0	435	490	462	470	481
	All	1,097	470.1	10.0	435	490	464	470	478
8	Autism	433	479.0	9.1	450	500	474	478	486
	Intellectual Disability - Mild	184	486.1	6.7	468	500	482	487	492
	Intellectual Disability - Moderate	306	480.9	7.6	450	500	476	481	486
	Multiple Disability	139	475.4	10.7	450	500	472	476	482
	Other	164	479.2	10.6	450	500	474	478	487
	All	1,226	480.2	9.3	450	500	475	480	487
10	Autism	288	489.7	9.6	460	510	485	489	495
	Intellectual Disability - Mild	127	495.9	8.2	460	510	492	497	502
	Intellectual Disability - Moderate	246	490.1	8.2	460	508	485	489	496
	Multiple Disability	90	484.5	12.4	460	504	481	486	493
	Other	121	490.8	10.7	460	508	485	490	501
	All	872	490.3	9.9	460	510	485	490	497

Table 8. 3 Science at Grades 5 and 8 and Biology NCEXTEND1 Scale Score by Accommodation Subgroups, Spring 2021

Grade/ Course	Students with Disability Category	N	Statistics		Range		Percentile		
			Average	SD	Min	Max	25th	Median	75th
5	Autism	457	450.6	9.8	415	470	444	450	458
	Intellectual Disability - Mild	167	456.3	6.7	436	470	452	457	461
	Intellectual Disability - Moderate	253	448.6	7.9	421	466	443	448	454
	Multiple Disability	128	444.8	11.1	415	466	440	446	452
	Other	129	450.0	11.9	415	470	443	451	460
	All	1,134	450.3	9.9	415	470	444	450	458
8	Autism	426	479.1	9.8	445	500	473	480	486
	Intellectual Disability - Mild	182	486.3	6.4	467	496	482	488	491
	Intellectual Disability - Moderate	304	480.3	8.3	445	500	475	480	487
	Multiple Disability	138	473.8	12.1	445	500	471	475	481
	Other	162	478.4	11.8	445	500	473	478	487
	All	1,212	479.8	10.2	445	500	474	480	487
Biology	Autism	288	489.4	10.7	455	510	484	490	497
	Intellectual Disability - Mild	127	496.0	7.4	455	510	492	497	501
	Intellectual Disability - Moderate	247	490.2	8.2	455	510	485	490	496
	Multiple Disability	92	483.0	13.8	455	504	481	486	492
	Other	120	489.6	11.3	455	510	483	490	498
	All	874	489.9	10.6	455	510	484	491	497

These results show that scale score distributions for some disability sub-groups from NCEXTEND1 administration have lower than the expected mean and standard deviation, specifically the one with smaller sample sizes. 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.

8.1.2 Scale Scores by Gender

Table 8.4 shows NCEXTEND1 scale score descriptive summary statistics for the Reading at grades 3–5 by gender, *Table 8.4* shows for grades 6–8 and 10, and *Table 8.5* shows for Science at grades 5 and 8 and Biology. Across grade levels, a higher proportion of male students (66%–69%) took NCEXTEND1 Reading and science tests in North Carolina during 2020–21 school year. In grades 3, 6, and 10 male students on average obtained higher mean scale score than female. In other grades, female students on average obtained slightly higher than male. The difference, however, is one (1) scale score point or less.

For science, male, on average, scored higher than female across grades with the difference ranging from 0.5 to 1.5. The largest difference was in grade 8.

Table 8. 4 Reading at Grades 3–5 NCEXTEND1 Scale Score by Gender, Spring 2021

Grade	Gender	N	Statistics		Range		Percentile		
			Average	SD	Min	Max	25th	Median	75th
3	Female	349	430.9	10.6	400	450	425	431	439
	Male	729	429.8	10.7	400	450	424	430	438
	All	1,078	430.1	10.7	400	450	424	430	439
4	Female	338	438.9	10.6	410	460	433	439	446
	Male	760	439.8	9.5	410	460	434	439	447
	All	1,098	439.5	9.8	410	460	434	439	446
5	Female	383	449.8	8.7	415	470	444	449	456
	Male	781	450.6	9.5	415	470	445	450	457
	All	1,164	450.3	9.2	415	470	445	450	457

Table 8. 5 Reading at Grades 6–8 and 10 NCEXTEND1 Scale Score by Gender, Spring 2021

Grade/ Course	Gender	N	Statistics		Range		Percentile		
			Average	SD	Min	Max	25th	Median	75th
6	Female	380	460.9	10.66	425	480	455	461	469
	Male	730	460.1	10.35	425	480	454	459	468
	All	1,110	460.4	10.46	425	480	454	460	468
7	Female	347	469.5	10.39	435	490	463	469	477
	Male	750	470.4	9.83	435	490	464	470	478
	All	1,097	470.1	10.01	435	490	464	470	478
8	Female	421	479.8	8.58	450	500	475	479	486
	Male	805	480.3	9.66	450	500	475	480	488
	All	1,226	480.2	9.3	450	500	475	480	487
10	Female	295	490.4	9.97	460	508	485	491	498
	Male	577	490.3	9.9	460	510	485	490	497
	All	872	490.3	9.92	460	510	485	490	497

Table 8. 6 Science at Grades 5 and 8 and Biology NCEXTEND1 Scale Score by Gender, Spring 2021

Grade/ Course	Gender	N	Statistics		Range		Percentile		
			Average	SD	Min	Max	25th	Median	75th
5	Female	374	449.7	9.25	415	470	444	450	457
	Male	760	450.6	10.22	415	470	444	451	459
	All	1,134	450.3	9.92	415	470	444	450	458
8	Female	414	478.8	9.39	445	500	473	479	485
	Male	798	480.3	10.51	445	500	474	481	488
	All	1,212	479.8	10.16	445	500	474	480	487
Biology	Female	297	489.6	10.64	455	510	484	491	496
	Male	577	490.1	10.6	455	510	484	490	498
	All	874	489.9	10.61	455	510	484	491	497

8.1.3 Scale Score by Major Ethnic Groups

For the purpose of this report, the scale scores descriptive statistics for the NCEXTEND1 assessments are summarized only for students who self-reported to be belonged to one of the three major ethnic groups: Black, Hispanic, and White. Students not self-identified in any of

those three major groups are classified as Other. *Table 8.7* to *Table 8.9* show the breakdown of NCEXTEND1 Reading and science scale scores by the ethnic groups from 2020–21 administration. The distribution of North Carolina alternate student population is very similar across grade levels with White students representing about 36% – 45% of all students across all grades and Black students representing about 31% – 36% with Hispanic students making about 16% – 22%. 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 8.7* to *Table 8.9* 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 Reading and science 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. 7 Reading at Grades 3–5 NCEXTEND1 Scale Score by Ethnicity, Spring 2021

Grade	Ethnic	N	Statistics		Range		Percentile		
			Average	SD	Min	Max	25th	Median	75th
3	Black	383	430.1	11.1	400	450	424	431	439
	Hispanic	179	430.2	10.4	400	450	425	430	438
	Other	112	429.4	11.45	400	450	424	430	439
	White	404	430.4	10.29	400	450	425	430	439
	All	1,078	430.1	10.71	400	450	424	430	439
4	Black	349	440.4	9.75	410	460	435	440	448
	Hispanic	239	438.9	9.16	410	460	433	438	445
	Other	118	437.8	9.92	410	460	433	438	444
	White	392	439.7	10.23	410	460	434	439	446
	All	1,098	439.5	9.84	410	460	434	439	446
5	Black	389	451	9.57	415	470	445	451	458
	Hispanic	214	449.8	9.02	415	470	444	450	455
	Other	120	451.2	7.66	436	470	446	450	457
	White	441	449.7	9.38	415	470	445	449	456
	All	1,164	450.3	9.23	415	470	445	450	457

Table 8. 8 Reading at Grades 6–8 and 10 NCEXTEND1 Scale Score by Ethnicity, Spring 2021

Grade/ Course	Ethnic	N	Statistics		Range		Percentile		
			Average	SD	Min	Max	25th	Median	75th
6	Black	385	460.7	10.33	425	480	455	460	469
	Hispanic	211	458.8	9.58	425	480	453	457	465
	Other	106	460.4	10.84	425	480	454	459	470
	White	408	460.9	10.86	425	480	454	461	469
	All	1,110	460.4	10.46	425	480	454	460	468
7	Black	371	470.6	10.06	435	490	464	471	479
	Hispanic	197	469.7	9.05	435	490	464	469	477
	Other	92	469.8	10.93	435	490	464	470	478
	White	437	469.9	10.19	435	490	464	469	477
	All	1,097	470.1	10.01	435	490	464	470	478
8	Black	400	480.8	9.41	450	500	475	481	488
	Hispanic	193	478.2	9.04	450	500	474	477	485
	Other	107	479.2	11.1	450	500	474	478	487
	White	526	480.6	8.82	450	500	475	481	487
	All	1,226	480.2	9.3	450	500	475	480	487
10	Black	273	490.2	9.65	460	510	485	490	497
	Hispanic	138	488.8	9.19	460	508	484	488	495
	Other	70	491.1	9.39	460	508	485	490	499
	White	391	490.8	10.41	460	510	486	492	498
	All	872	490.3	9.92	460	510	485	490	497

Table 8. 9 Science at Grades 5 and 8 and Biology NCEXTEND1 Scale Score by Ethnicity, Spring 2021

Grade/ Course	Ethnic	N	Statistics		Range		Percentile		
			Average	SD	Min	Max	25th	Median	75th
5	Black	378	451.3	9.8	415	470	445	452	460
	Hispanic	207	449.0	9.9	415	470	443	449	457
	Other	116	450.9	9.0	426	470	444	450	459
	White	433	449.8	10.2	415	470	444	450	458
	All	1,134	450.3	9.9	415	470	444	450	458
8	Black	394	480.1	10.3	445	500	474	481	487
	Hispanic	190	478.3	9.6	445	500	472	478	485
	Other	107	478.2	13.1	445	500	473	480	487
	White	521	480.4	9.5	445	500	474	481	487
	All	1,212	479.8	10.2	445	500	474	480	487
Biology	Black	275	489.5	10.8	455	510	484	490	497
	Hispanic	138	489.2	9.7	455	510	483	490	496
	Other	71	490.6	10.8	455	510	483	492	499
	White	390	490.4	10.8	455	510	485	491	498
	All	874	489.9	10.6	455	510	484	491	497

8.1.4 Achievement Levels Distributions

Beginning in 2020–21 with *Edition 4* NCEXTEND1 tests, the NCDPI transitioned to classify and report student performance on the Reading and science 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 Reading and science 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 8.11 and *Figure 8.12* show summary of proportion of students by achievement level from the 2020–21 NCEXTEND1 Reading and science assessments respectively. The stacked bar graphs show classifications by grade. For example, 59% students in grade 3 (*Figure 8.11*) are classified as Not Proficient, 30% Level 3, and 11% Level 4 indicating 41% (Level 3 and above) of NC grade 3 students who took the NCEXTEND1 Reading assessments are considered to have met grade level content expectations for state accountability reporting purposes. While about 11% 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. 11 State Level NCEXTEND1 Reading at Grades 3–8 and 10 Achievement Level Classifications, Spring 2021

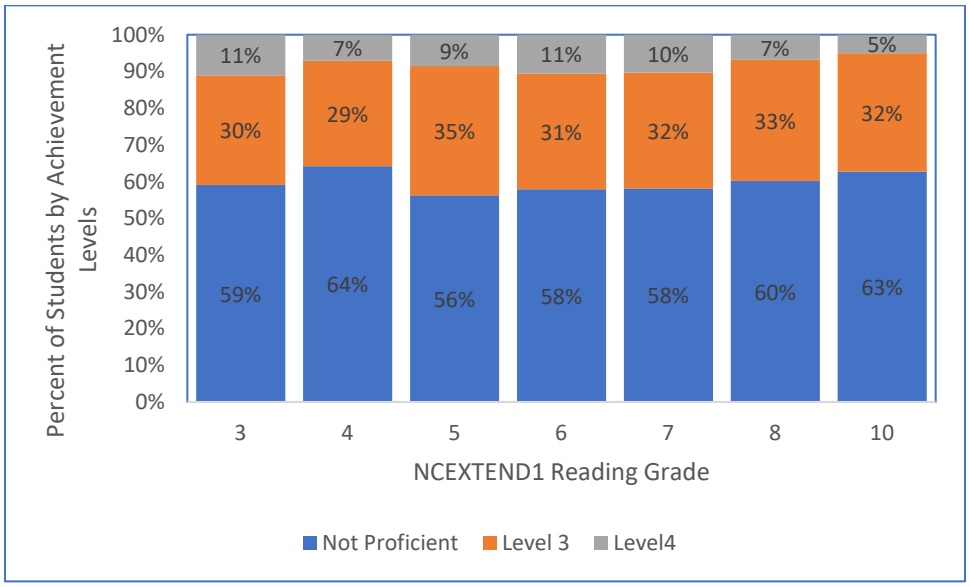
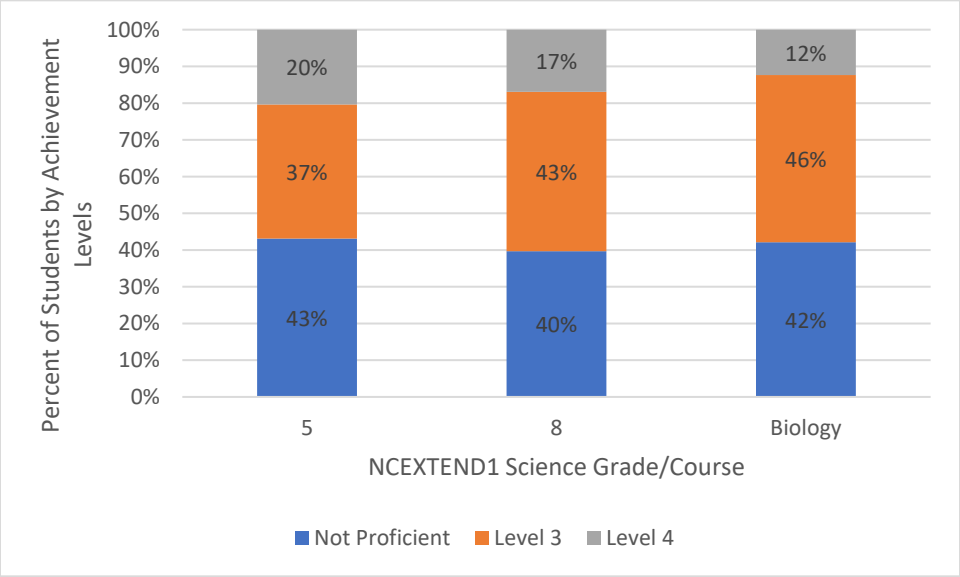


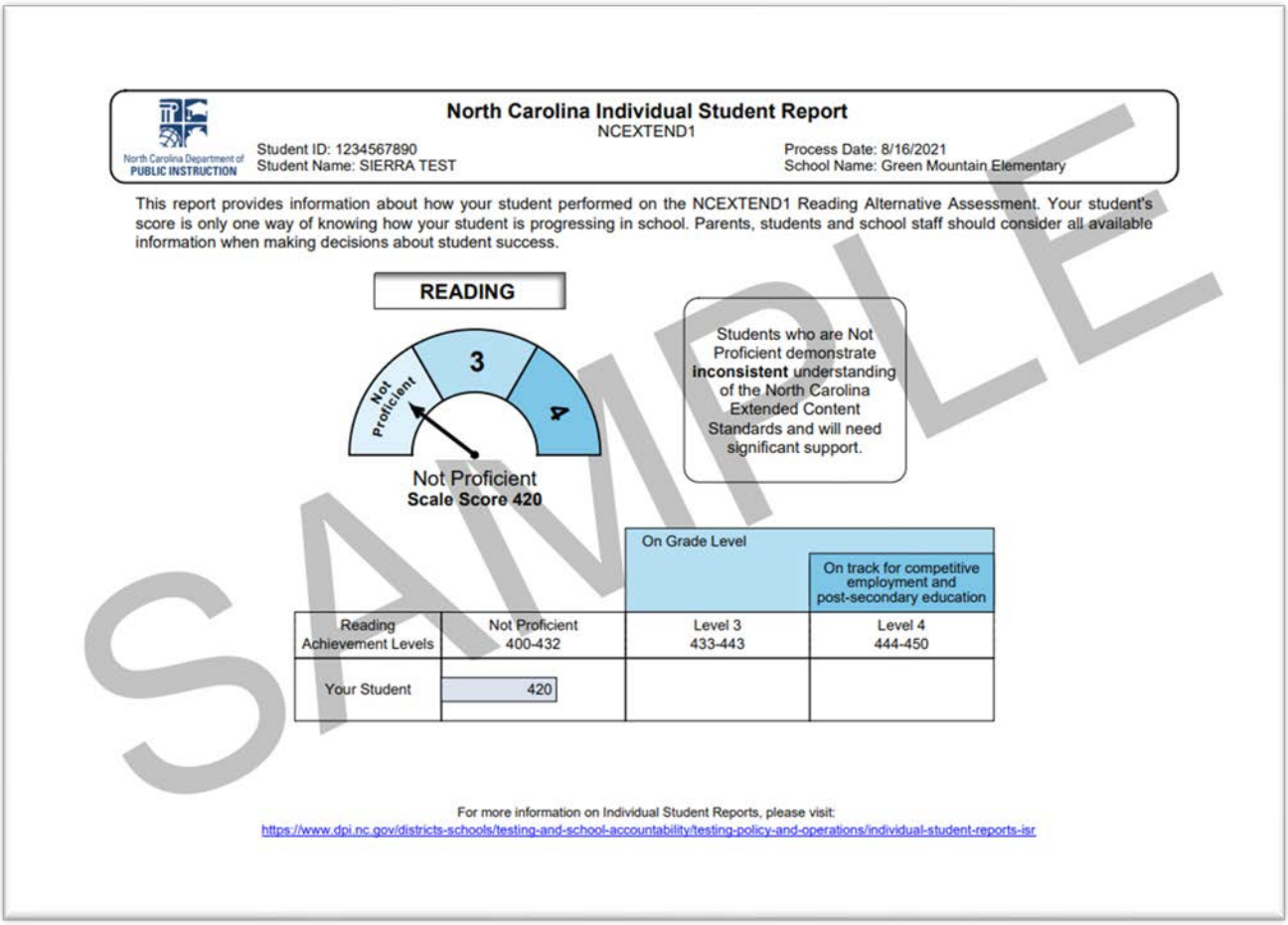
Figure 8. 12 State Level NCEXTEND1 Science at Grades 5 and 8 and Biology Achievement Level Classifications, Spring 2021



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) template shown in *Figure 8.13* 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/uisrs.

Figure 8. 13 *Individual Student Report (ISR)*



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 8.10* shows a summary list of the main pre-programmed static reports that are currently available to the different audiences for NCEXTEND1 Reading 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. 10 Reports by Audience

WinScan Report	Audience			
	Teacher	School	District	State
Class Roster Reports	✓	✓		
Score and Achievement Level Frequency	✓	✓	✓	✓
Goal Summary Reports	✓	✓	✓	✓

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 EVIDENCE

This chapter presents additional validity evidence collected in support of the interpretation of *Edition 4* grades 3–8 and 10 Reading and science grades 5 and 8 and Biology 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 Reading and science scores. The final section of the chapter documents procedures used to ensure *Edition 4* NCEXTEND1 alternate assessment 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:

$$\hat{\alpha} = \frac{\kappa}{\kappa-1}\left(1 - \frac{\Sigma\hat{\sigma}_i^2}{\hat{\sigma}_X^2}\right) \qquad (9\text{–}1)$$

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 Reading and Science NCEXTEND1 alternate assessments by grade and major demographic variables for 2020–21 administration. Across grades, overall reliability estimates based on the 2020–21 population ranged from the lowest of 0.86 to the highest of 0.90. Subgroup reliabilities are also in the similar range with the lowest generally 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 Grades 3–8 Reading, English II at Grade 10, Grades 5 and 8 Science, and Biology at Grade 10

Grade/ Course	Gender		Ethnicity[1]			Accommodations[1]		Overall	SE
	Female	Male	Black	Hispanic	White	EDS	ELs		
Reading									
3	0.90	0.89	0.90	0.89	0.88	0.89	0.88	0.90	3.47
4	0.89	0.86	0.87	0.84	0.88	0.88	0.82	0.87	3.54
5	0.83	0.86	0.87	0.84	0.85	0.86	0.80	0.86	3.50
6	0.89	0.88	0.88	0.86	0.89	0.89	0.83	0.89	3.52
7	0.89	0.87	0.88	0.85	0.88	0.88	0.83	0.88	3.49
8	0.82	0.87	0.86	0.83	0.84	0.86	0.80	0.86	3.49
10	0.87	0.87	0.86	0.85	0.88	0.86	0.83	0.87	3.52
Science									
5	0.85	0.88	0.88	0.87	0.88	0.88	0.84	0.88	3.47
8	0.85	0.89	0.88	0.85	0.86	0.88	0.85	0.88	3.50
Biology	0.89	0.89	0.90	0.86	0.89	0.88	0.85	0.89	3.45

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

9.2 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 adequate yearly progress (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-Class[2]. 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 Reading and science NCEXTEND1 alternate assessment followed the HB procedures.

Table 9.2 shows the decision accuracy and consistency indexes for the NCEXTEND1 alternate assessment' achievement cuts at each grade. Two observations can be made from the Table for Reading: 1) the decision accuracy at both levels is 0.89 or higher, and 2) consistency values are higher (0.88–0.94) for the Level 4 compared to Level 3 (0.85 – 0.88).

For Science, the decision accuracy ranged from 0.83 to 0.93. Decision consistency was higher for Level 3 compared to Level 4 for all grades/level.

The grade 3 results in *Table 9.2* can be interpreted as if grade 3 NCEXTEND1 Reading students who were classified as Level 3 were to take a non-overlapping, equally difficult form a second time, 88% (**bolded**) of them would still be classified as Level 3 and 88% (**bolded**) of them would still be classified as Level 4. The classification accuracy for the grade 3 Reading Level 3 indicated that there is 93% agreement between the observed and true classification.

[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.

Table 9. 2 NCEXTEND1 Reading and Science Classification Accuracy and Consistency Results

Grade/ Course	Level 3		Level 4	
	Sufficient Understanding		Thorough Understanding	
	Grade Level Proficient		Competitive Employment	
	Accuracy	Consistency	Accuracy	Consistency
Reading				
3	0.92	**0.88**	0.89	**0.88**
4	0.90	0.86	0.93	0.92
5	0.90	0.85	0.91	0.90
6	0.91	0.88	0.89	0.89
7	0.90	0.87	0.90	0.89
8	0.90	0.85	0.93	0.92
10	0.90	0.86	0.95	0.94
Science				
5	0.91	0.88	0.89	0.86
8	0.93	0.90	0.83	0.81
Biology	0.93	0.90	0.87	0.85

9.3 Unidimensionality of NCEXTEND1 Assessments

The North Carolina Reading and science NCEXTEND1 alternate assessment are designed based on a unidimensional assumption that total score represents an estimate of students' performance based on grade level 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 Reading and science 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 9.10* show the PCA results for all Reading and Science NCEXTEND alternate assessment forms. The left vertical axis shows the actual eigenvalues of a test 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 Reading and Science NCEXTEND1 alternate assessments.

Figure 9. 1 Grade 3 NCEXTEND1 Reading Scree Plot of 2020–21 Operational Form

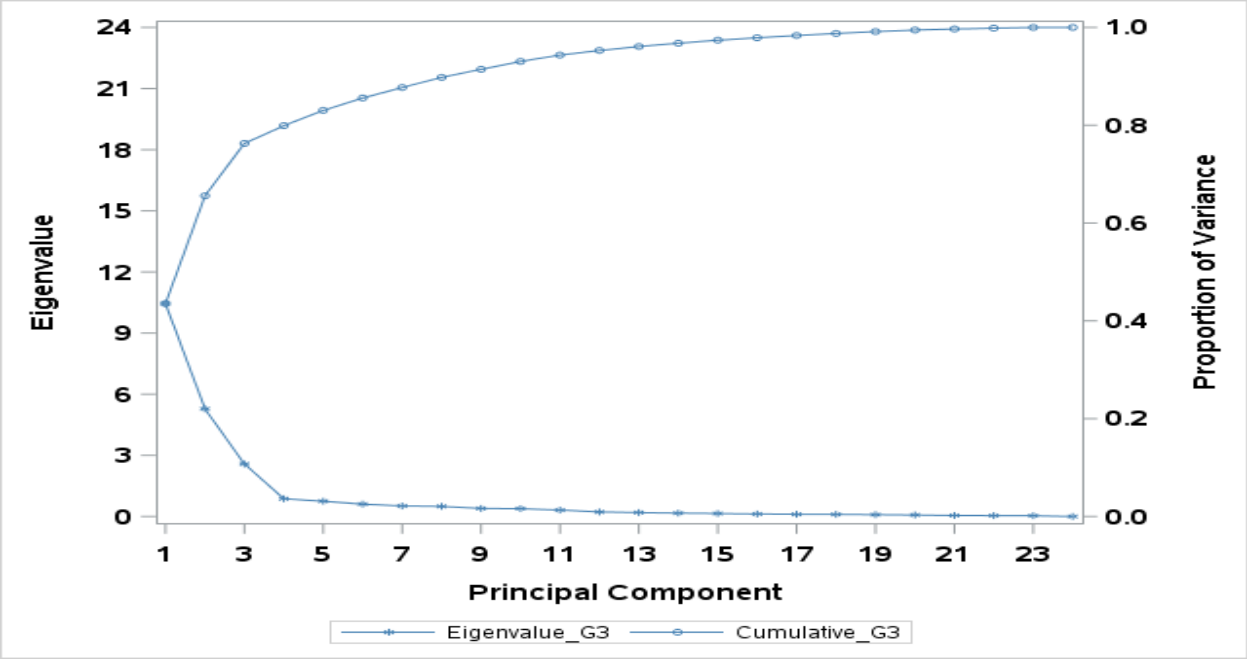


Figure 9. 2 Grade 4 NCEXTEND1 Reading Scree Plot of 2020–21 Operational Form

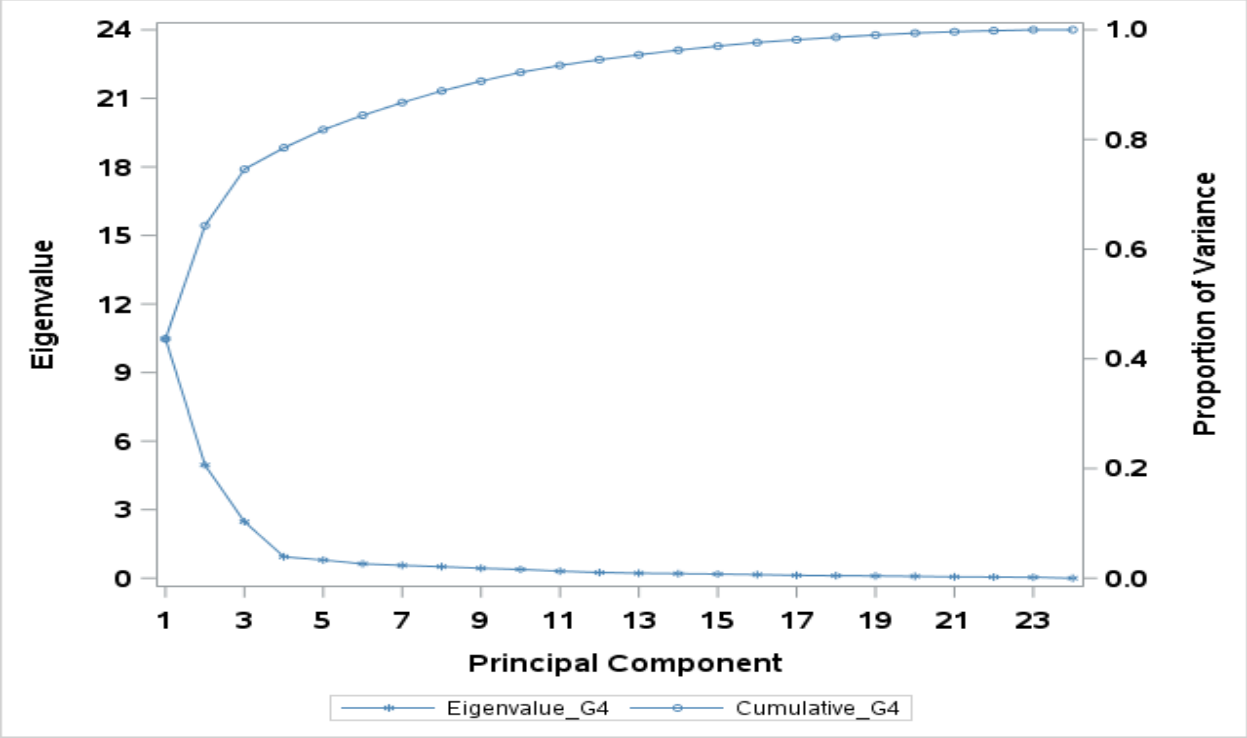


Figure 9. 3 Grade 5 NCEXTEND1 Reading Scree Plot of 2020–21 Operational Form

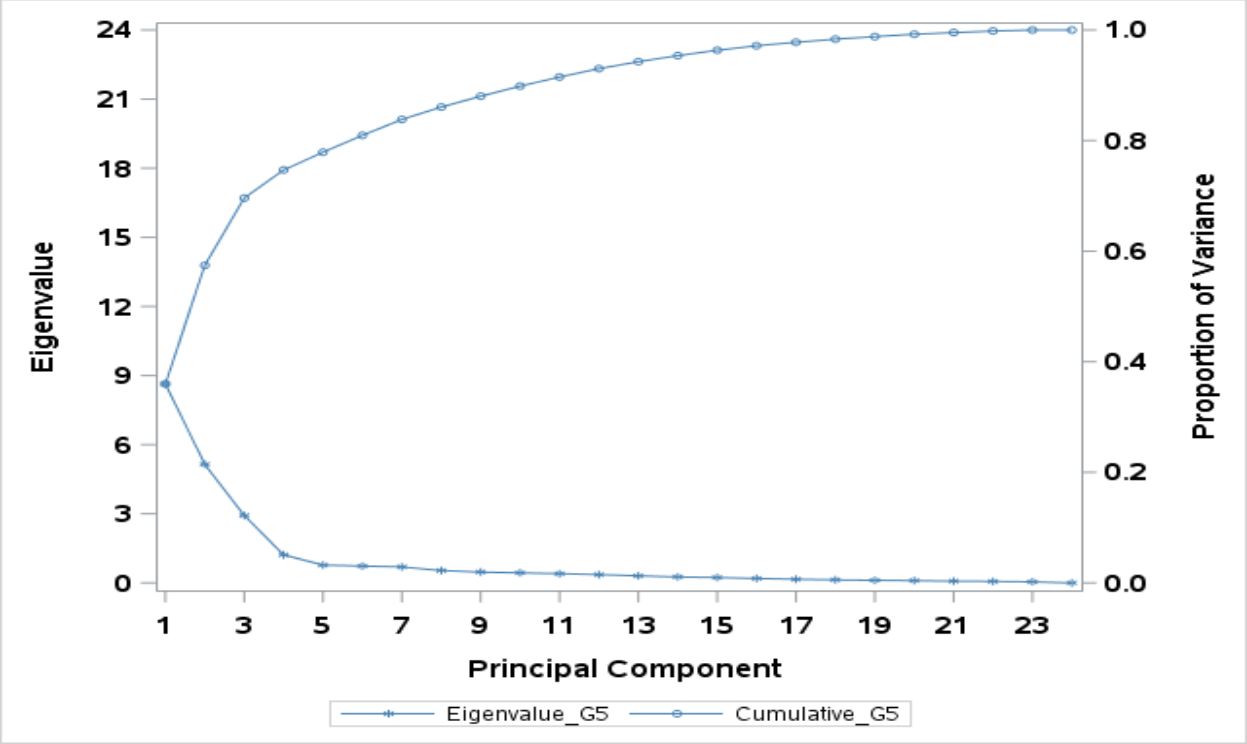


Figure 9. 4 Grade 6 NCEXTEND1 Reading Scree Plot of 2020–21 Operational Form

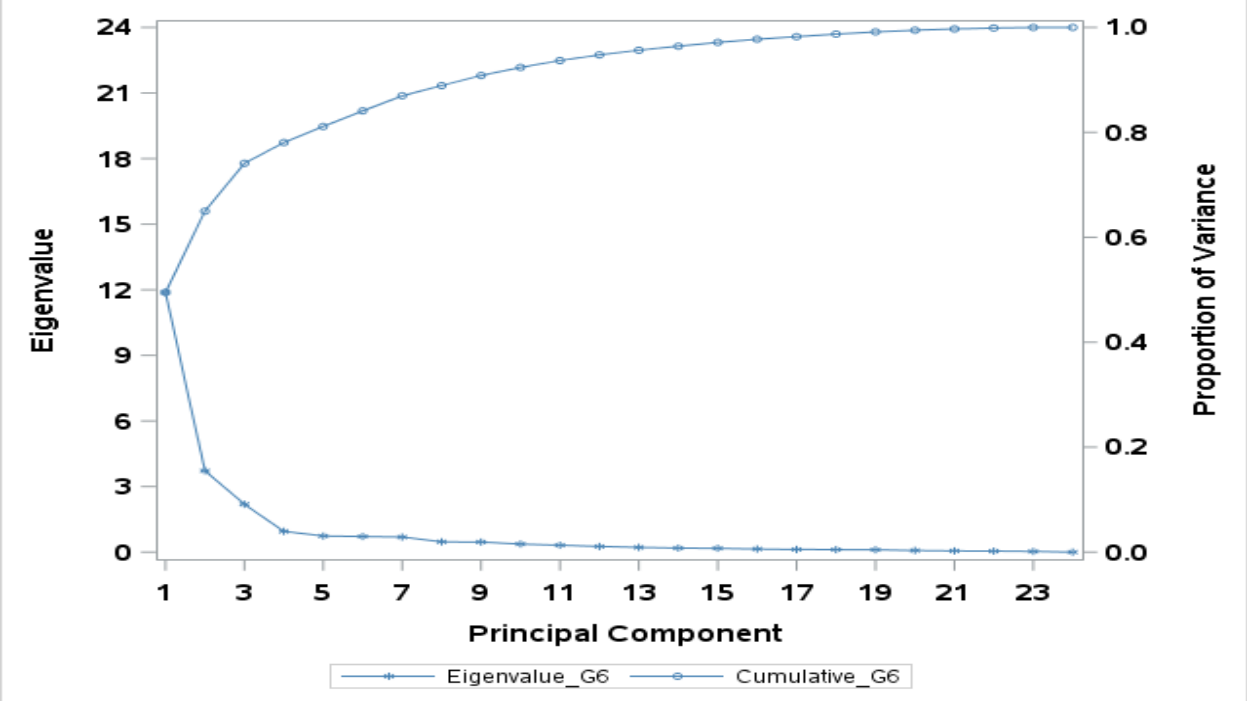


Figure 9. 5 Grade 7 NCEXTEND1 Reading Scree Plot of 2020–21 Operational Form

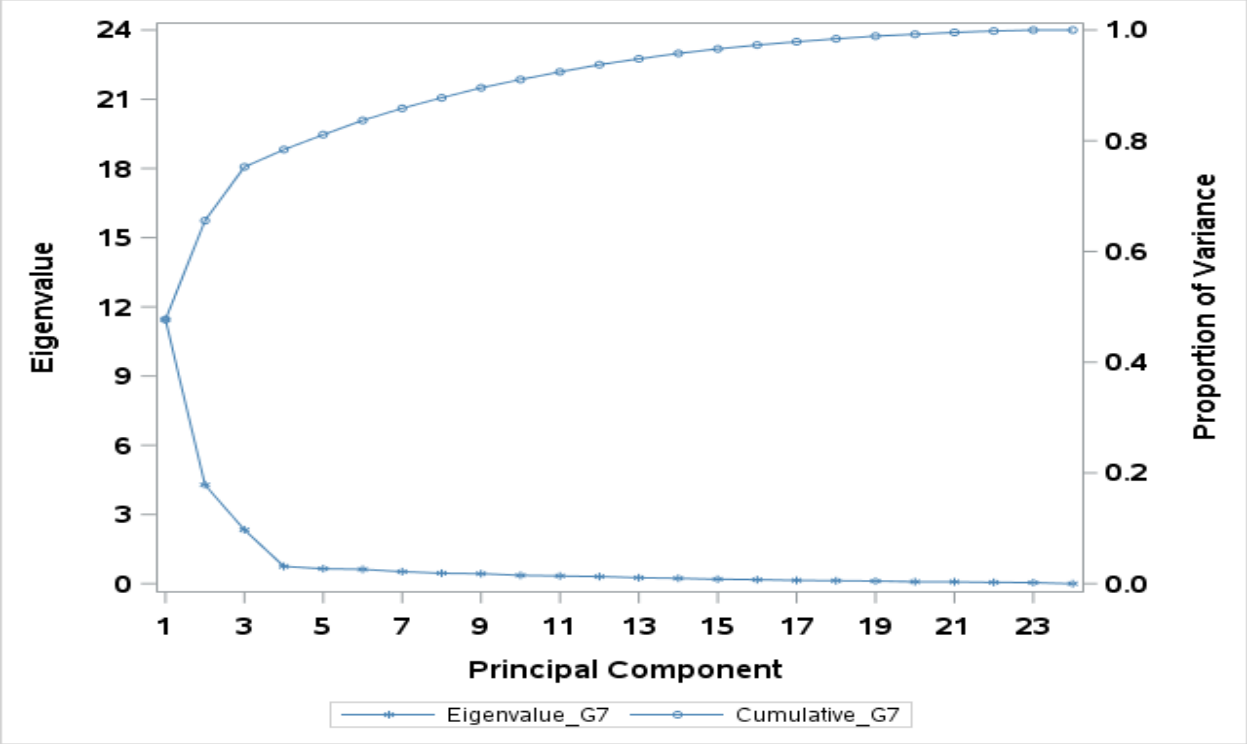


Figure 9. 6 Grade 8 NCEXTEND1 Reading Scree Plot of 2020–21 Operational Form

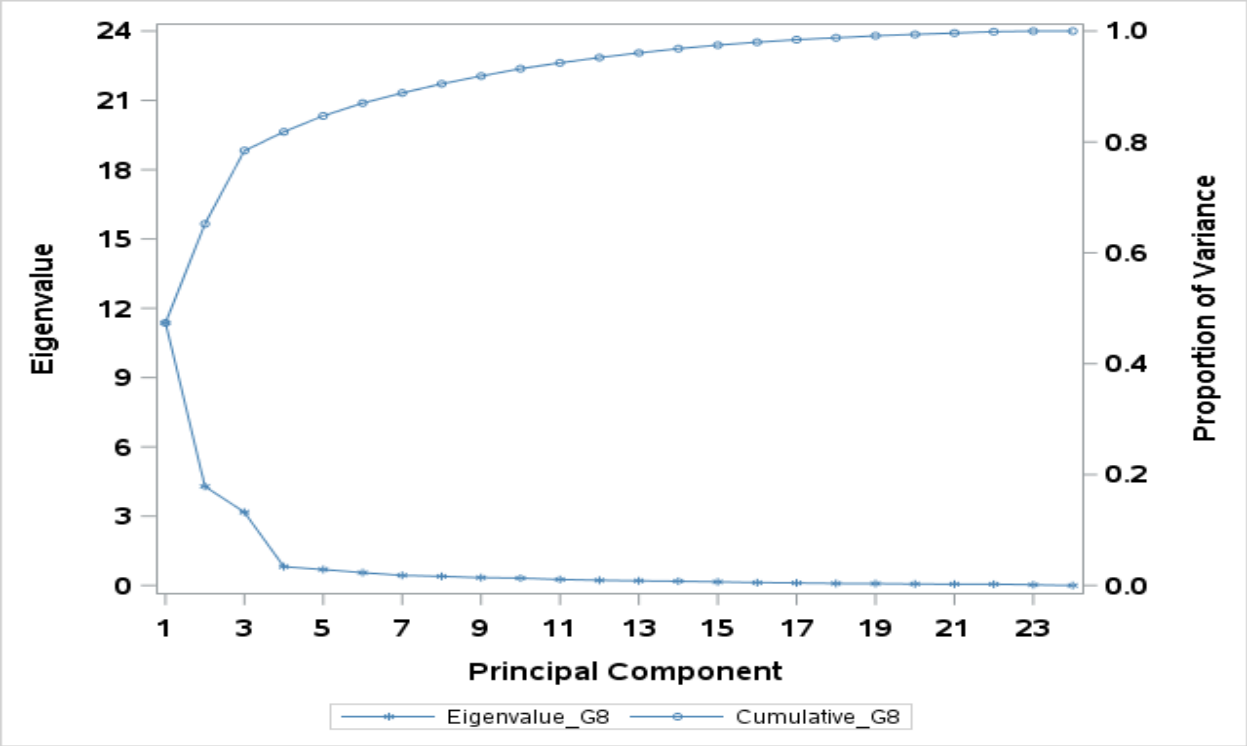


Figure 9. 7 NCEXTEND1 English II at Grade 10 Scree Plot of 2020–21 Operational Form

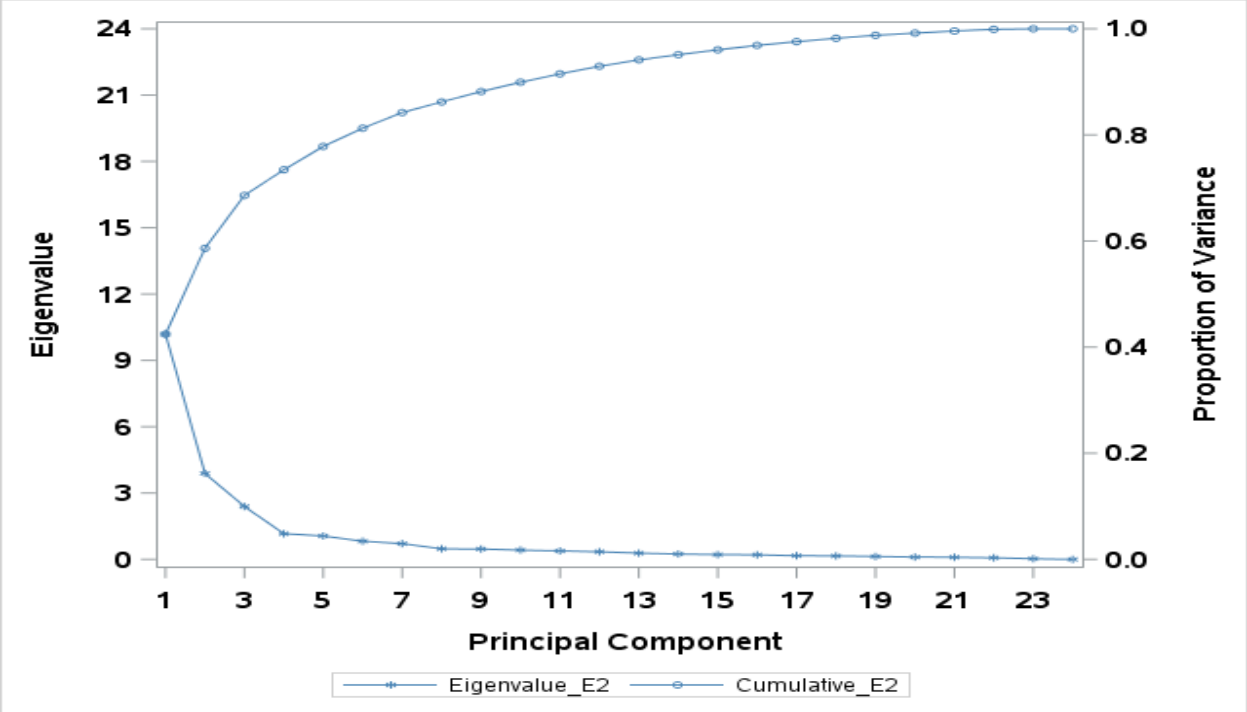


Figure 9. 8 Grade 5 NCEXTEND1 Science Scree Plot of 2020–21 Operational Form

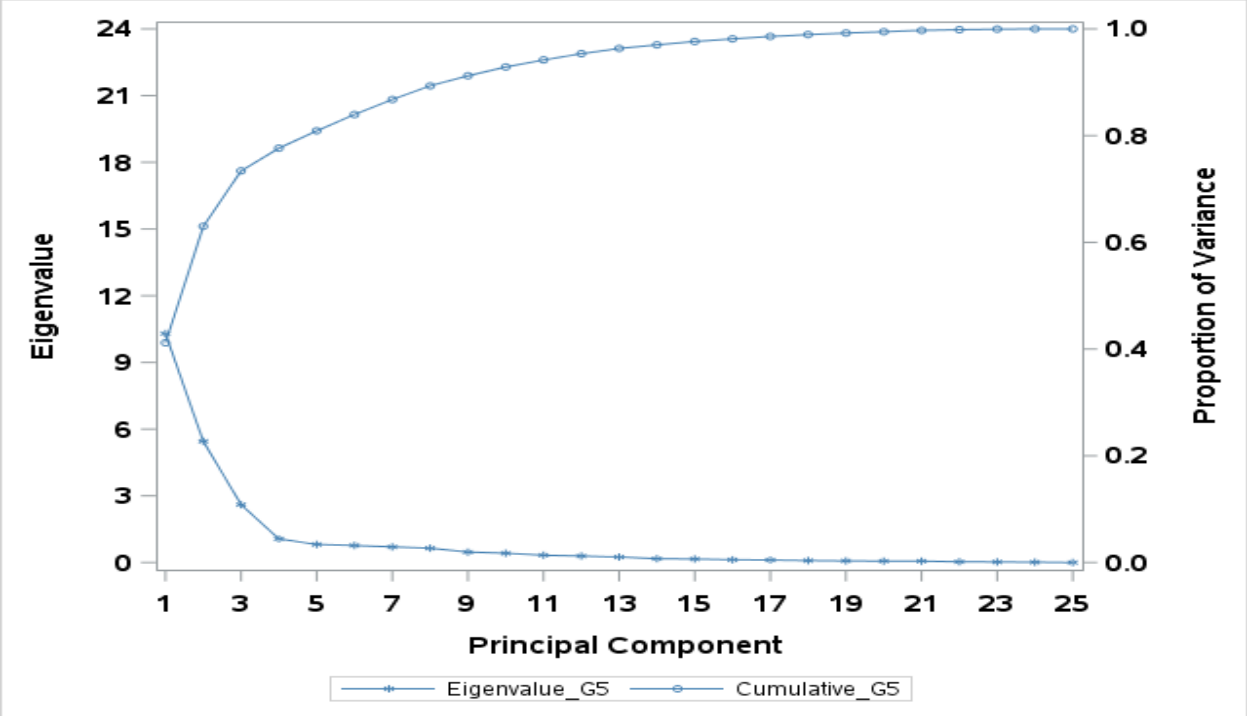


Figure 9. 9 Grade 8 NCEXTEND1 Science Scree Plots of 2020–21 Operational Forms

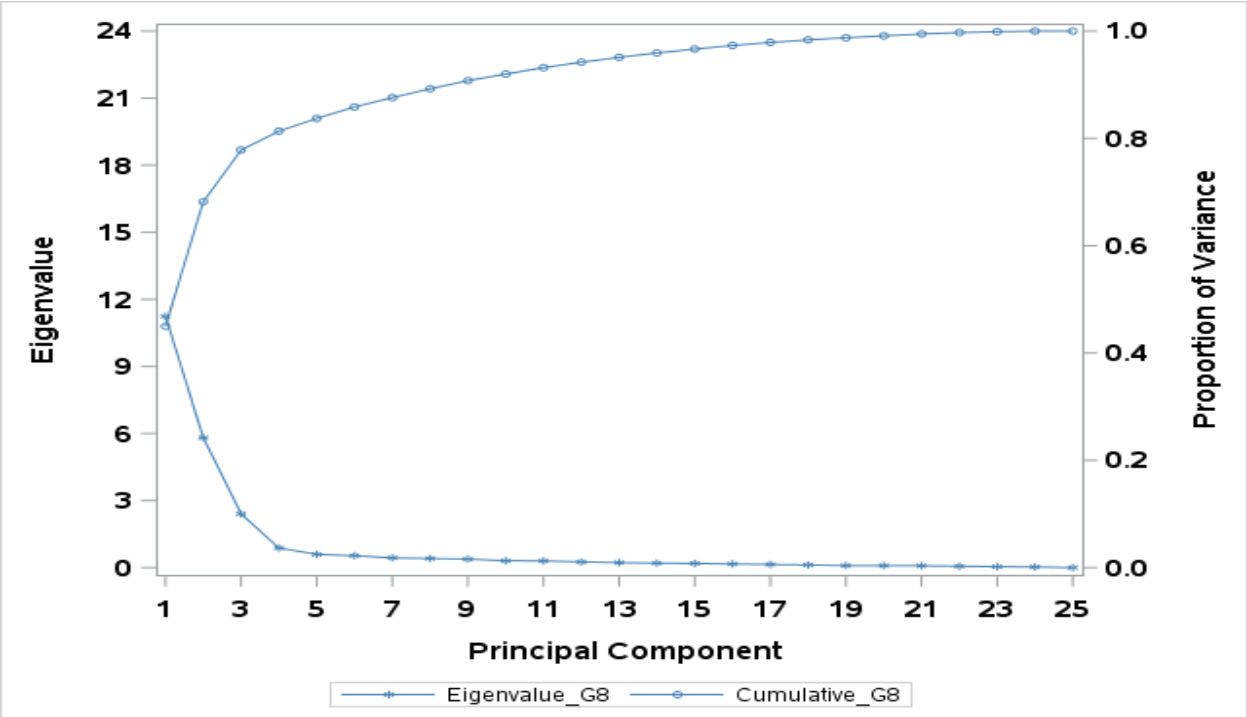
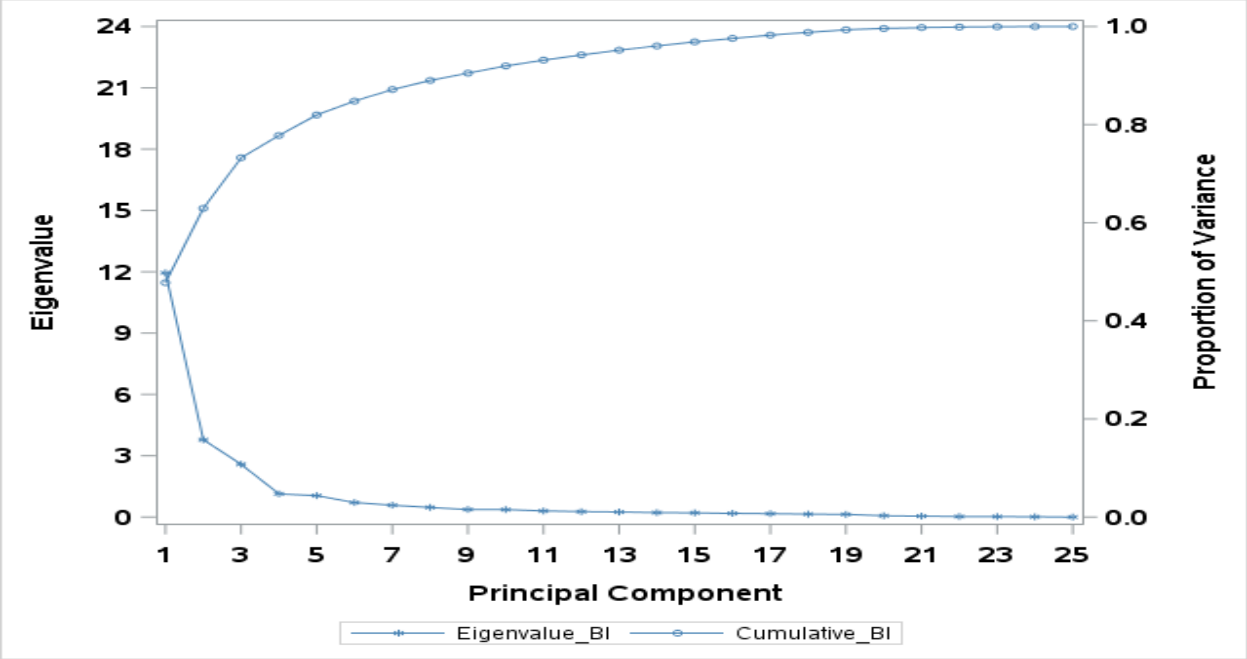


Figure 9. 10 NCEXTEND1 Biology Scree Plot of 2020–21 Operational Form



The Eigen values and variance accounted for by the components with Eigenvalues greater than one are summarized in *Table 9.4* for Reading and *Table 9.5* for science. Based on the PCA results, the average ratio of the first to the second eigenvalue across grades for Reading ranged from a minimum of approximately 2.0 in grades 3–5 to a maximum of 3.0 in grades 6–8 and 10. For science, the average ratio of the first to second eigenvalue across grades for Science ranged from a minimum of approximately 1.9 in grades 5 and 8 to 3.2 in grade 10. These results indicate significantly smaller proportion of variance explained by the second factor. Moreover, on average the first principal component accounted for about the lowest of 36% in grade 5 Reading to the highest of 49% in grade 6 Reading.

Table 9. 3 Eigenvalues (Eigen) and Variance (%) Accounted for by the Components, Reading at Grades 3–8, and English II at Grade 10

Factor	Grade/Course													
	3		4		5		6		7		8		10	
	Eigen	%	Eigen	%	Eigen	%	Eigen	%	Eigen	%	Eigen	%	Eigen	%
1	10.5	44	10.5	44	8.6	36	11.9	49	11.5	48	11.4	47	10.2	42
2	5.3	22	5.0	21	5.1	21	3.7	16	4.3	18	4.3	18	3.9	16
3	2.6	11	2.5	10	2.9	12	2.2	9	2.3	10	3.2	13	2.4	10
4					1.2	5							1.2	5
5													1.1	4

Table 9. 4 Eigenvalues (Eigen) and Variance (%) Accounted for by the Components, Science at Grades 5 and 8 and Biology

Factor	Grade/Course					
	5		8		Biology	
	Eigen	%	Eigen	%	Eigen	%
1	10.3	41%	11.2	45%	11.9	48%
2	5.5	22%	5.8	23%	3.8	15%
3	2.6	10%	2.4	10%	2.6	10%
4	1.1	4%			1.1	5%
5					1.0	4%

9.4 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 conducted in 2022-23 school year. The report is available in the NCDPI website.

9.5 Fairness and Accessibility

9.5.1 Accessibility in Universal Design

To ensure fairness and accessibility for all eligible students for the grades 3–8 and 10 Reading and grades 5, 8, and Biology NCEXTEND1 alternate assessment, 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 items[3]
    - Accommodations included from the start (Braille, large–print, oral presentation etc.)
    - Ensuring that quality is retained in all items
- Simple, clear directions and procedures
    - 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.

9.5.2 Fairness in Access

[3] See discussions on fairness review in Chapter 4

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 assessment. 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, the NCDPI also published released items for every grade level and content, 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.

9.5.3 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 Reading 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 ona 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

Testing Code of Ethics (nc.gov)

Appendix 2–A

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

English Language Arts
Test Specifications Meeting Agenda
February 26, 2018
Jane S. McKimmon Center, NC State University

8:30am	**Registration—Main Lobby**
9:00am	**Welcome and Introductions** Dr. Tammy Howard, Dan Auman • Meeting purpose • Substitute Teacher Form, Stipend Form, Demographics Form • Testing Code of Ethics and Test Security Agreement • Travel Reimbursement
9:30am	**Summative Assessment Psychometric Overview** Dr. Kinge Mbella
10:15am	**Break**
10:30am	**Overview of Revised ELA Standards** DPI-Curriculum & Instruction and Exceptional Children Divisions
11:30am	**Prioritizing Standards Overview** Dan Auman
11:45am	**Lunch** (on your own)
12:45pm	**Prioritize Standards—ROUND 1** (Breakout Groups—General and EC: Grades 3-5, Grades 6-8, and Grades 9-12) • Prioritize Assessable Standards • Recommend Weighting by Domain
2:15pm	**Break** (on your own)
2:30pm	**Prioritize Standards—ROUND 2** (Breakout Groups) • Prioritize Assessable Standards • Recommend Weighting by Domain
3:15pm	**Recommend Percent by Item Type—Discussion** (Large Group) Dan Auman, Kinge Mbella
3:45pm	**Summary of Recommendations and General Considerations** Dan Auman
4:00 pm	**Meeting Adjourned**

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

Appendix 2–B

Reading Depth-of-Knowledge Levels

Preliminary Depth of Knowledge Levels (nciea.org)

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

open (nc.gov)

Appendix 3–A

Fairness and DIF Review Process

Appendix 5–A

Testing Students with disability Handbook

Testing Students with disability Handbook 2021 (nc.gov)

Appendix 5–B

Testing Security Protocols and Procedures for School Personnel

Testing Security Protocols and Procedures for School Personnel 2021 (nc.gov)

Appendix 5–C

North Carolina Test Coordinators' Policies and Procedures Handbook

Test Coordinator Policies and Procedures Handbook 2022 (nc.gov)

Appendix 5–D

North Carolina Alternate Assessment Eligibility Criteria

NCEXTEND1_eligibility_criteria_2019.pdf

Appendix 5-E

Request for Testing Exceptions Based on Significant Medical Emergencies and/or Conditions

Med Exception Memo (nc.gov).

Appendix 5-F

Online Testing Irregularity Submission System (OTISS) User Manual

otissmanual11 Updated Mar 2019 (nc.gov)

Appendix 7–A

Standard Setting Technical Report: Edition 4 NCEXTEND1 Mathematics 2021

Technical Information for State Tests | NC DPI

Appendix 7–B

Observation Report: Edition 4 NCEXTEND1 Reading Standard Setting 2020-21

Technical Information for State Tests | NC DPI

Appendix 8–A

Subgroups Distribution: NCEXTEND1 Reading Scale Score Descriptive Statistics

Table 1. 2020-21 NCEXTEND1 Grade 3 Reading Scale Score Descriptive Statistics by Subgroups

Group	Categories	N	Statistics		Range		Percentile		
			Average	SD	Min	Max	25th	Median	75th
Ethnicity	Asian	33	426.6	10.6	400	444	423	426	434
	Black	373	430.2	11.1	400	450	424	431	439
	Hispanic	171	430.2	10.3	400	450	425	430	438
	American Indian	17	434.0	12.1	400	446	424	436	444
	Multiracial	57	429.9	11.1	401	450	424	430	439
	Native Hawaiian/Pacific Islander	3	431.3	4.7	426	435	426	433	435
	White	391	430.4	10.3	400	450	425	429	439
	All	1,045	430.2	10.7	400	450	424	430	439
SWD	Autism	501	429.0	10.7	400	450	424	429	436
	Deaf-Blindness	1	427.0		427	427	427	427	427
	Serious Emotional Disability	3	440.0	10.4	428	447	428	445	447
	Hearing Impairment	1	441.0		441	441	441	441	441
	Intellectual Disability - Mild	128	436.9	7.3	420	450	431	439	443
	Intellectual Disability - Moderate	181	431.7	9.0	400	450	426	431	439
	Intellectual Disability - Severe	33	422.4	9.6	400	441	420	424	428
	Specific Learning Disability	6	440.3	6.7	430	446	434	444	445
	Multiple Disability	116	425.3	11.5	400	447	422	427	432
	Other Health Impairment	52	435.0	9.3	412	450	428	436	444
	Orthopedic Impairment	1	401.0		401	401	401	401	401
	Other	33	428.2	12.1	400	446	424	429	438
	Traumatic Brain Injury	21	429.5	10.7	403	447	426	431	436
	Visual Impairment	1	412.0		412	412	412	412	412
	All	1078	430.1	10.7	400	450	424	430	439
EDS	Not Economically Disadvantaged	507	429.0	10.8	400	450	424	429	436
	Economically Disadvantaged	538	431.4	10.4	400	450	425	431	440
	All	1045	430.2	10.7	400	450	424	430	439
Els	Regular	926	430.1	10.8	400	450	424	430	439
	Other	7	431.6	8.0	424	445	425	429	437
	English Language Learner	112	430.6	9.6	400	447	425	430	439
	All	1045	430.2	10.7	400	450	424	430	439

Table 2. 2020-21 NCEXTEND1 Grade 4 Reading Scale Score Descriptive Statistics by Subgroups

Group	Categories	N	Statistics		Range		Percentile		
			Average	SD	Min	Max	25th	Median	75th
Ethnicity	Asian	45	435.7	10.2	410	454	432	437	440
	Black	339	440.6	9.7	410	460	435	440	448
	Hispanic	233	438.9	9.2	410	460	433	438	445
	American Indian	14	439.5	11.9	410	460	432	441	447
	Multiracial	55	439.2	9.3	410	455	435	439	447
	Native Hawaiian/Pacific Islander	2	434.0	11.3	426	442	426	434	442
	White	384	439.7	10.2	410	460	434	439	446
	All	1072	439.6	9.8	410	460	434	439	447
SWD	Autism	471	438.7	9.9	410	460	433	439	445
	Deafness	1	438.0		438	438	438	438	438
	Serious Emotional Disability	1	456.0		456	456	456	456	456
	Intellectual Disability - Mild	143	445.8	8.1	420	460	439	447	452
	Intellectual Disability - Moderate	242	439.8	7.3	410	460	435	438	444
	Intellectual Disability - Severe	44	433.2	9.0	410	446	430	436	438
	Specific Learning Disability	6	452.7	1.8	450	455	452	453	454
	Multiple Disability	116	435.1	11.7	410	457	430	436	443
	Other Health Impairment	40	445.2	8.1	429	460	441	444	453
	Other	26	436.4	9.4	410	451	434	438	441
	Traumatic Brain Injury	8	436.3	13.4	410	453	432	436	447
	All	1098	439.5	9.8	410	460	434	439	446
EDS	Not Economically Disadvantaged	556	439.1	9.6	410	460	433	439	445
	Economically Disadvantaged	516	440.1	10.0	410	460	435	440	448
	All	1072	439.6	9.8	410	460	434	439	447
Els	Regular	919	439.6	10.1	410	460	434	439	447
	Other	2	441.0	7.1	436	446	436	441	446
	English Language Learner	151	439.5	8.5	410	457	433	438	445
	All	1072	439.6	9.8	410	460	434	439	447

Table 3. 2020-21 NCEXTEND1 Grade 5 Reading Scale Score Descriptive Statistics by Subgroups

Group	Categories	N	Statistics		Range		Percentile		
			Average	SD	Min	Max	25th	Median	75th
Ethnicity	Asian	45	449.7	7.6	436	466	444	449	456
	Black	384	451.0	9.6	415	470	445	451	458
	Hispanic	211	449.9	8.8	415	470	444	450	455
	American Indian	16	453.3	7.3	442	465	449	452	460
	Multiracial	54	452.6	7.6	440	470	446	452	459
	Native Hawaiian/Pacific Islander	4	442.8	5.1	436	448	439	444	447
	White	434	449.7	9.4	415	470	445	449	456
	All	1,148	450.3	9.2	415	470	445	450	457
SWD	Autism	475	450.4	9.5	415	470	445	450	457
	Deaf-Blindness	2	442.5	3.5	440	445	440	443	445
	Serious Emotional Disability	4	463.0	2.5	460	465	461	464	465
	Hearing Impairment	1	456.0		456	456	456	456	456
	Intellectual Disability - Mild	174	455.6	6.4	441	470	451	456	460
	Intellectual Disability - Moderate	256	449.1	7.0	415	466	444	449	453
	Intellectual Disability - Severe	40	443.6	10.9	415	466	441	446	450
	Specific Learning Disability	8	460.0	8.5	445	470	455	462	466
	Multiple Disability	128	445.6	9.8	415	466	441	447	451
	Other Health Impairment	43	453.3	9.3	415	470	449	453	461
	Orthopedic Impairment	7	449.7	6.7	440	460	446	447	455
	Other	16	449.4	11.8	421	466	442	450	459
	Traumatic Brain Injury	9	447.6	12.7	424	464	442	445	459
	Visual Impairment	1	453.0		453	453	453	453	453
	All	1164	450.3	9.2	415	470	445	450	457
EDS	Not Economically Disadvantaged	586	449.9	9.1	415	470	445	449	456
	Economically Disadvantaged	562	450.8	9.3	415	470	446	451	458
	All	1148	450.3	9.2	415	470	445	450	457
Els	Regular	981	450.5	9.4	415	470	445	450	457
	Other	3	447.0	1.7	446	449	446	446	449
	English Language Learner	164	449.3	8.3	415	470	444	450	455
	All	1148	450.3	9.2	415	470	445	450	457

Table 4. 2020-21 NCEXTEND1 Grade 6 Reading Scale Score Descriptive Statistics by Subgroups

Group	Categories	N	Statistics		Range		Percentile		
			Average	SD	Min	Max	25th	Median	75th
Ethnicity	Asian	36	458.4	9.4	430	476	453	458	466
	Black	379	460.7	10.4	425	480	455	460	469
	Hispanic	204	458.7	9.6	425	480	453	457	465
	American Indian	15	460.5	13.2	425	480	455	459	470
	Multiracial	52	461.8	11.2	425	480	455	461	471
	White	405	460.9	10.9	425	480	454	461	469
	All	1091	460.4	10.5	425	480	454	460	468
SWD	Autism	428	459.1	10.8	425	480	453	458	466
	Deaf-Blindness	2	443.5	7.8	438	449	438	444	449
	Deafness	2	457.0	0.0	457	457	457	457	457
	Serious Emotional Disability	3	473.7	2.3	471	475	471	475	475
	Hearing Impairment	1	463.0		463	463	463	463	463
	Intellectual Disability - Mild	162	465.8	8.4	425	480	460	467	472
	Intellectual Disability - Moderate	267	460.6	8.1	431	480	455	459	467
	Intellectual Disability - Severe	42	452.5	9.2	425	476	450	453	457
	Specific Learning Disability	6	470.7	2.9	466	473	469	472	473
	Multiple Disability	106	455.5	12.8	425	480	452	456	462
	Other Health Impairment	52	466.6	9.9	444	480	460	469	475
	Orthopedic Impairment	4	465.3	6.0	460	473	461	464	470
	Other	19	463.2	8.4	447	475	457	461	471
	Traumatic Brain Injury	15	459.9	8.5	444	476	456	459	463
	Visual Impairment	1	474.0		474	474	474	474	474
	All	1110	460.4	10.5	425	480	454	460	468
EDS	Not Economically Disadvantaged	542	459.7	10.4	425	480	453	459	468
	Economically Disadvantaged	549	461.0	10.6	425	480	455	461	469
	All	1091	460.4	10.5	425	480	454	460	468
Els	Regular	941	460.4	10.8	425	480	454	460	469
	Other	2	465.5	6.4	461	470	461	466	470
	English Language Learner	148	459.8	8.6	430	480	454	458	466
	All	1091	460.4	10.5	425	480	454	460	468

Table 5. 2020-21 NCEXTEND1 Grade 7 Reading Scale Score Descriptive Statistics by Subgroups

Group	Categories	N	Statistics		Range		Percentile		
			Average	SD	Min	Max	25th	Median	75th
Ethnicity	Asian	21	469.0	6.6	455	483	466	468	472
	Black	368	470.6	10.1	435	490	464	471	479
	Hispanic	193	469.7	9.1	435	490	464	469	477
	American Indian	9	475.1	9.6	461	490	467	477	482
	Multiracial	58	469.6	12.4	435	490	463	470	480
	Native Hawaiian/Pacific Islander	1	461.0		461	461	461	461	461
	White	431	469.9	10.2	435	490	464	469	477
	All	1081	470.1	10.0	435	490	464	470	478
SWD	Autism	380	468.1	9.8	435	490	463	468	475
	Serious Emotional Disability	10	482.2	3.3	476	486	481	483	485
	Hearing Impairment	3	468.3	6.1	463	475	463	467	475
	Intellectual Disability - Mild	182	477.1	6.7	459	490	472	478	482
	Intellectual Disability - Moderate	289	469.8	8.6	435	490	463	470	476
	Intellectual Disability - Severe	41	459.2	11.9	435	484	458	461	465
	Specific Learning Disability	9	478.9	5.9	467	485	478	479	484
	Multiple Disability	96	466.3	8.6	435	486	463	466	470
	Other Health Impairment	50	472.8	13.1	435	490	465	476	483
	Orthopedic Impairment	2	464.5	5.0	461	468	461	465	468
	Other	16	471.1	9.3	460	485	462	470	480
	Traumatic Brain Injury	17	472.9	9.4	455	490	467	470	479
	Visual Impairment	2	481.5	0.7	481	482	481	482	482
	All	1097	470.1	10.0	435	490	464	470	478
EDS	Not Economically Disadvantaged	554	469.1	9.9	435	490	463	468	476
	Economically Disadvantaged	527	471.2	10.1	435	490	465	472	479
	All	1081	470.1	10.0	435	490	464	470	478
Els	Regular	941	470.1	10.3	435	490	464	470	478
	Other	3	467.3	3.8	463	470	463	469	470
	English Language Learner	137	470.1	8.3	441	490	464	470	476
	All	1081	470.1	10.0	435	490	464	470	478

Table 6. 2020-21 NCEXTEND1 Grade 8 Reading Scale Score Descriptive Statistics by Subgroups

Group	Categories	N	Statistics	Range	Percentile

			Average	SD	Min	Max	25th	Median	75th
Ethnicity	Asian	36	477.0	8.7	450	495	474	477	481
	Black	392	480.8	9.5	450	500	475	481	488
	Hispanic	189	478.2	9.1	450	500	474	477	485
	American Indian	10	483.5	15.3	450	497	474	489	496
	Multiracial	56	480.6	11.2	450	500	475	480	489
	Native Hawaiian/Pacific Islander	2	474.5	5.0	471	478	471	475	478
	White	520	480.6	8.8	450	500	475	481	487
	All	1205	480.2	9.3	450	500	475	480	487
SWD	Autism	433	479.0	9.1	450	500	474	478	486
	Deaf-Blindness	2	481.0	7.1	476	486	476	481	486
	Deafness	2	472.0	4.2	469	475	469	472	475
	Serious Emotional Disability	2	475.0	12.7	466	484	466	475	484
	Hearing Impairment	1	475.0		475	475	475	475	475
	Intellectual Disability - Mild	184	486.1	6.7	468	500	482	487	492
	Intellectual Disability - Moderate	306	480.9	7.6	450	500	476	481	486
	Intellectual Disability - Severe	51	472.9	8.7	450	487	471	474	478
	Specific Learning Disability	1	488.0		488	488	488	488	488
	Multiple Disability	139	475.4	10.7	450	500	472	476	482
	Other Health Impairment	62	483.7	10.1	450	500	477	484	492
	Orthopedic Impairment	1	492.0		492	492	492	492	492
	Other	21	479.1	9.1	450	490	474	479	486
	Traumatic Brain Injury	18	480.1	11.1	453	496	475	478	492
	Visual Impairment	3	490.3	9.1	482	500	482	489	500
	All	1226	480.2	9.3	450	500	475	480	487
EDS	Not Economically Disadvantaged	640	479.7	9.2	450	500	474	479	486
	Economically Disadvantaged	565	480.8	9.4	450	500	475	481	488
	All	1205	480.2	9.3	450	500	475	480	487
Els	Regular	1082	480.3	9.4	450	500	475	480	487
	Other	1	475.0		475	475	475	475	475
	English Language Learner	122	479.3	8.4	450	500	475	479	485
	All	1205	480.2	9.3	450	500	475	480	487

Table 7. 2020-21 NCEXTEND1 English II at Grade 10 Scale Score Descriptive Statistics by Subgroups

Group	Categories	N	Statistics		Range		Percentile		
			Average	SD	Min	Max	25th	Median	75th
Ethnic	Asian	15	490.6	11.1	460	508	487	490	498
	Black	273	490.2	9.7	460	510	485	490	497
	Hispanic	138	488.8	9.2	460	508	484	488	495
	American Indian	14	492.4	9.3	476	507	485	491	499
	Multiracial	40	490.7	8.9	460	506	485	490	498
	Native Hawaiian/Pacific Islander	1	501.0		501	501	501	501	501
	White	391	490.8	10.4	460	510	486	492	498
	All	872	490.3	9.9	460	510	485	490	497
SWD	Autism	288	489.7	9.6	460	510	485	489	495
	Deaf-Blindness	2	493.0	12.7	484	502	484	493	502
	Serious Emotional Disability	2	491.5	16.3	480	503	480	492	503
	Intellectual Disability - Mild	127	495.9	8.2	460	510	492	497	502
	Intellectual Disability - Moderate	246	490.1	8.2	460	508	485	489	496
	Intellectual Disability - Severe	45	485.0	9.1	460	506	481	486	489
	Specific Learning Disability	5	504.2	1.6	503	507	503	504	504
	Multiple Disability	90	484.5	12.4	460	504	481	486	493
	Other Health Impairment	44	492.8	10.5	460	508	486	493	502
	Orthopedic Impairment	2	489.0	18.4	476	502	476	489	502
	Other	10	494.0	10.1	469	504	491	495	501
	Traumatic Brain Injury	10	496.2	7.6	482	503	488	500	501
	Visual Impairment	1	501.0		501	501	501	501	501
	All	872	490.3	9.9	460	510	485	490	497
EDS	Not Economically Disadvantaged	520	490.0	10.3	460	510	485	490	497
	Economically Disadvantaged	352	490.8	9.3	460	510	485	491	497
	All	872	490.3	9.9	460	510	485	490	497
ELS	Regular	785	490.4	10.1	460	510	485	490	497
	Other	4	482.3	6.6	474	488	477	484	488
	English Language Learner	83	490.1	8.7	460	508	484	488	496
	All	872	490.3	9.9	460	510	485	490	497

Table 8. 2020-21 NCEXTEND1 Grade 5 Science Scale Score Descriptive Statistics by Subgroups

Group	Categories	N	Statistics	Range	Percentile

			Average	SD	Min	Max	25th	Median	75th
Ethnicity	Asian	44	449.1	9.2	431	470	442	447	455
	Black	378	451.3	9.8	415	470	445	452	460
	Hispanic	207	449.0	9.9	415	470	443	449	457
	American Indian	15	453.8	8.0	441	466	448	452	463
	Multiracial	53	452.2	8.4	438	466	444	455	459
	Native Hawaiian/Pacific Islander	4	441.8	10.9	426	451	435	445	449
	White	433	449.8	10.2	415	470	444	450	458
	All	1,134	450.3	9.9	415	470	444	450	458
SWD	Autism	457	450.6	9.8	415	470	444	450	458
	Deaf-Blindness	2	438.5	2.1	437	440	437	439	440
	Serious Emotional Disability	4	464.3	4.0	461	470	462	463	467
	Hearing Impairment	1	456.0		456	456	456	456	456
	Intellectual Disability - Mild	167	456.3	6.7	436	470	452	457	461
	Intellectual Disability - Moderate	253	448.6	7.9	421	466	443	448	454
	Intellectual Disability - Severe	40	443.4	10.8	415	465	438	442	450
	Specific Learning Disability	8	458.6	6.9	447	470	455	459	463
	Multiple Disability	128	444.8	11.1	415	466	440	446	452
	Other Health Impairment	42	454.6	10.0	415	466	448	457	462
	Orthopedic Impairment	7	452.0	4.9	446	458	446	452	457
	Other	15	448.3	14.6	418	466	441	452	461
	Traumatic Brain Injury	9	447.3	13.5	421	464	443	445	461
	Visual Impairment	1	450.0		450	450	450	450	450
	All	1134	450.3	9.9	415	470	444	450	458
EDS	Not Economically Disadvantaged	583	449.6	9.8	415	470	444	450	457
	Economically Disadvantaged	551	451.0	10.0	415	470	445	452	459
	All	1134	450.3	9.9	415	470	444	450	458
Els	Regular	974	450.5	10.0	415	470	444	451	459
	Other	3	443.3	6.4	436	448	436	446	448
	English Language Learner	157	449.0	9.2	415	466	443	449	456
	All	1134	450.3	9.9	415	470	444	450	458

Table 9. 2020-21 NCEXTEND1 Grade 8 Science Scale Score Descriptive Statistics by Subgroups

Group	Categories	N	Statistics	Range	Percentile

			Average	SD	Min	Max	25th	Median	75th
Ethnicity	Asian	36	474.7	12.2	445	494	472	475	482
	Black	394	480.1	10.3	445	500	474	481	487
	Hispanic	190	478.3	9.6	445	500	472	478	485
	American Indian	10	484.2	15.9	445	496	482	489	496
	Multiracial	58	480.1	12.1	445	500	474	482	488
	Native Hawaiian/Pacific Islander	3	463.3	15.9	445	473	445	472	473
	White	521	480.4	9.5	445	500	474	481	487
	All	1212	479.8	10.2	445	500	474	480	487
SWD	Autism	426	479.1	9.8	445	500	473	480	486
	Deaf-Blindness	1	483.0		483	483	483	483	483
	Deafness	2	472.5	3.5	470	475	470	473	475
	Serious Emotional Disability	2	464.5	27.6	445	484	445	465	484
	Hearing Impairment	1	476.0		476	476	476	476	476
	Intellectual Disability - Mild	182	486.3	6.4	467	496	482	488	491
	Intellectual Disability - Moderate	304	480.3	8.3	445	500	475	480	487
	Intellectual Disability - Severe	51	471.6	10.1	445	488	470	473	478
	Specific Learning Disability	1	488.0		488	488	488	488	488
	Multiple Disability	138	473.8	12.1	445	500	471	475	481
	Other Health Impairment	62	483.1	10.7	445	500	478	484	491
	Orthopedic Impairment	1	487.0		487	487	487	487	487
	Other	20	478.1	10.9	445	493	473	478	488
	Traumatic Brain Injury	18	480.5	11.6	451	500	474	483	487
	Visual Impairment	3	490.7	9.0	482	500	482	490	500
	All	1212	479.8	10.2	445	500	474	480	487
EDS	Not Economically Disadvantaged	640	479.3	10.1	445	500	473	480	487
	Economically Disadvantaged	572	480.3	10.2	445	500	474	481	488
	All	1212	479.8	10.2	445	500	474	480	487
Els	Regular	1089	479.9	10.2	445	500	474	481	487
	Other	1	471.0		471	471	471	471	471
	English Language Learner	122	478.6	9.6	445	500	473	479	484
	All	1212	479.8	10.2	445	500	474	480	487

Table 10. 2020-21 NCEXTEND1 Biology Grade 10 Scale Score Descriptive Statistics by Subgroups

Group	Categories	N	Statistics		Range		Percentile		
			Average	SD	Min	Max	25th	Median	75th
Ethnic	Asian	15	488.8	13.9	455	510	479	488	498
	Black	275	489.5	10.8	455	510	484	490	497
	Hispanic	138	489.2	9.7	455	510	483	490	496
	American Indian	14	493.4	10.4	473	506	485	496	503
	Multiracial	41	490.1	9.6	455	504	483	490	498
	Native Hawaiian/Pacific Islander	1	503.0		503	503	503	503	503
	White	390	490.4	10.8	455	510	485	491	498
	All	874	489.9	10.6	455	510	484	491	497
SWD	Autism	288	489.4	10.7	455	510	484	490	497
	Deaf-Blindness	2	494.5	16.3	483	506	483	495	506
	Serious Emotional Disability	2	492.0	1.4	491	493	491	492	493
	Intellectual Disability - Mild	127	496.0	7.4	455	510	492	497	501
	Intellectual Disability - Moderate	247	490.2	8.2	455	510	485	490	496
	Intellectual Disability - Severe	46	484.5	9.9	455	505	481	484	488
	Specific Learning Disability	4	505.8	2.9	504	510	504	505	508
	Multiple Disability	92	483.0	13.8	455	504	481	486	492
	Other Health Impairment	44	491.0	11.8	455	510	485	492	500
	Orthopedic Impairment	2	494.0	12.7	485	503	485	494	503
	Other	9	492.3	12.8	464	504	490	498	498
	Traumatic Brain Injury	10	494.4	5.0	486	501	490	496	499
	Visual Impairment	1	503.0		503	503	503	503	503
	All	874	489.9	10.6	455	510	484	491	497
EDS	Not Economically Disadvantaged	520	489.7	10.9	455	510	484	490	497
	Economically Disadvantaged	354	490.4	10.1	455	510	485	491	497
	All	874	489.9	10.6	455	510	484	491	497
ELS	Regular	785	489.9	10.8	455	510	484	491	497
	Other	4	486.3	10.4	477	498	478	485	495
	English Language Learner	85	490.7	9.2	455	510	484	491	496
	All	874	489.9	10.6	455	510	484	491	497

Appendix 8–B

Achievement Level Ranges and Descriptors

Reading (nc.gov)

Science (nc.gov)

Appendix 8–C

NCEXTEND1 Proficiency Classifications by Subgroups

Table 1. 2020-21 NCEXTEND1 Grade 3 Reading Proficiency Classifications by Subgroups

Category	Sub-Category	N	Not Proficient	Level 3	Level 4
Ethnicity	Asian	34	26.5	2.9	70.6
	Black	383	31.9	10.7	57.4
	Hispanic	179	26.8	11.2	62.0
	American Indian	17	29.4	29.4	41.2
	Multiracial	58	27.6	12.1	60.3
	Native Hawaiian/Pacific Islander	3	66.7		33.3
	White	404	29.7	11.1	59.2
	All	1078	29.9	11.0	59.1
SWD	Autism	501	27.0	9.0	64.1
	Deaf-Blindness	1			100.0
	Serious Emotional Disability	3		66.7	33.3
	Hearing Impairment	1	100.0		
	Intellectual Disability - Mild	128	49.2	21.1	29.7
	Intellectual Disability - Moderate	181	33.2	11.6	55.3
	Intellectual Disability - Severe	33	9.1		90.9
	Specific Learning Disability	6	33.3	50.0	16.7
	Multiple Disability	116	20.7	3.5	75.9
	Other Health Impairment	52	32.7	25.0	42.3
	Orthopedic Impairment	1			100.0
	Other	33	27.3	9.1	63.6
	Traumatic Brain Injury	21	38.1	4.8	57.1
	Visual Impairment	1			100.0
	All	1078	29.9	11.0	59.1
EDS	Not Economically Disadvantaged	519	27.9	8.7	63.4
	Economically Disadvantaged	559	31.7	13.2	55.1
	All	1078	29.9	11.0	59.1
ELs	Regular	952	30.0	11.0	58.9
	Other	9	22.2	11.1	66.7
	English Language Learner	117	29.1	11.1	59.8
	All	1078	29.9	11.0	59.1

Note: Not Proficient=Inconsistent Understanding, Not CCR 3= Sufficient Understanding, Not CCR 4= Thorough Understanding, on Track for Competitive Employment, CCR

Table 2. 2020-21 NCEXTEND1 Grade 4 Reading Proficiency Classifications by Subgroups

Category	Sub-Category	N	Not Proficient	Level 3	Level 4
Ethnicity	Asian	46	19.6	2.2	78.3
	Black	349	30.4	8.9	60.7
	Hispanic	239	27.2	5.4	67.4
	American Indian	14	35.7	7.1	57.1
	Multiracial	55	27.3	3.6	69.1
	Native Hawaiian/Pacific Islander	3			100.0
	White	392	29.3	7.7	63.0
	All	1098	28.7	7.1	64.2
SWD	Autism	471	25.5	6.4	68.2
	Deafness	1			100.0
	Serious Emotional Disability	1		100.0	
	Intellectual Disability - Mild	143	46.9	17.5	35.7
	Intellectual Disability - Moderate	242	30.2	2.9	66.9
	Intellectual Disability - Severe	44	11.4		88.6
	Specific Learning Disability	6	66.7	33.3	
	Multiple Disability	116	21.6	3.5	75.0
	Other Health Impairment	40	37.5	22.5	40.0
	Other	26	11.5		88.5
	Traumatic Brain Injury	8	37.5		62.5
	All	1098	28.7	7.1	64.2
EDS	Not Economically Disadvantaged	575	26.3	6.3	67.5
	Economically Disadvantaged	523	31.4	8.0	60.6
	All	1098	28.7	7.1	64.2
ELs	Regular	944	28.8	7.3	63.9
	Other	2	50.0		50.0
	English Language Learner	152	27.6	5.9	66.5
	All	1098	28.7	7.1	64.2

Note: Not Proficient=Inconsistent Understanding, Not CCR 3= Sufficient Understanding, Not CCR 4= Thorough Understanding, on Track for Competitive Employment, CCR

Table 3. 2020-21 NCEXTEND1 Grade 5 Reading Proficiency Classifications by Subgroups

Category	Sub-Category	N	Not Proficient	Level 3	Level 4
Ethnicity	Asian	45	24.4	8.9	66.7
	Black	389	39.1	9.3	51.7
	Hispanic	214	37.4	7.0	55.6
	American Indian	16	31.3	18.8	50.0
	Multiracial	55	38.2	10.9	50.9
	Native Hawaiian/Pacific Islander	4			100.0
	White	441	32.0	7.9	60.1
	All	1164	35.2	8.5	56.3
SWD	Autism	475	33.3	9.9	56.8
	Deaf-Blindness	2			100.0
	Serious Emotional Disability	4	50.0	50.0	
	Hearing Impairment	1	100.0		
	Intellectual Disability - Mild	174	59.8	14.4	25.9
	Intellectual Disability - Moderate	256	30.5	3.1	66.4
	Intellectual Disability - Severe	40	15.0	2.5	82.5
	Specific Learning Disability	8	25.0	50.0	25.0
	Multiple Disability	128	19.5	3.1	77.3
	Other Health Impairment	43	51.2	11.6	37.2
	Orthopedic Impairment	7	42.9		57.1
	Other	16	31.3	12.5	56.3
	Traumatic Brain Injury	9	33.3	11.1	55.6
	Visual Impairment	1	100.0		
	All	1164	35.2	8.5	56.3
EDS	Not Economically Disadvantaged	594	32.0	8.3	59.8
	Economically Disadvantaged	570	38.6	8.8	52.6
	All	1164	35.2	8.5	56.3
ELs	Regular	997	35.1	9.2	55.7
	Other	3			100.0
	English Language Learner	164	36.6	4.3	59.2
	All	1164	35.2	8.5	56.3

Note: Not Proficient=Inconsistent Understanding, Not CCR 3= Sufficient Understanding, Not CCR 4= Thorough Understanding, on Track for Competitive Employment, CCR

Table 4. 2020-21 NCEXTEND1 Grade 6 Reading Proficiency Classifications by Subgroups

Category	Sub-Category	N	Not Proficient	Level 3	Level 4
Ethnicity	Asian	37	21.6	8.1	70.3
	Black	385	35.6	9.4	55.1
	Hispanic	211	24.6	7.1	68.3
	American Indian	16	25.0	18.8	56.3
	Multiracial	53	22.6	20.8	56.6
	White	408	33.3	12.3	54.4
	All	1110	31.4	10.6	57.9
SWD	Autism	428	27.6	9.8	62.6
	Deaf-Blindness	2			100.0
	Deafness	2			100.0
	Serious Emotional Disability	3	33.3	66.7	
	Hearing Impairment	1	100.0		
	Intellectual Disability - Mild	162	48.8	18.5	32.7
	Intellectual Disability - Moderate	267	34.8	4.9	60.3
	Intellectual Disability - Severe	42	7.1	2.4	90.5
	Specific Learning Disability	6	100.0		
	Multiple Disability	106	17.9	6.6	75.5
	Other Health Impairment	52	34.6	32.7	32.7
	Orthopedic Impairment	4	50.0		50.0
	Other	19	26.3	15.8	57.9
	Traumatic Brain Injury	15	26.7	13.3	60.0
	Visual Impairment	1		100.0	
	All	1110	31.4	10.6	57.9
EDS	Not Economically Disadvantaged	554	28.0	9.8	62.3
	Economically Disadvantaged	556	34.9	11.5	53.6
	All	1110	31.4	10.6	57.9
ELs	Regular	958	31.7	11.3	57.0
	Other	2	50.0		50.0
	English Language Learner	150	29.3	6.7	64.0
	All	1110	31.4	10.6	57.9

Note: Not Proficient=Inconsistent Understanding, Not CCR 3= Sufficient Understanding, Not CCR 4= Thorough Understanding, on Track for Competitive Employment, CCR

Table 5. 2020-21 NCEXTEND1 Grade 7 Reading Proficiency Classifications by Subgroups

Category	Sub-Category	N	Not Proficient	Level 3	Level 4
Ethnicity	Asian	22	18.2	4.6	77.3
	Black	371	34.5	10.5	55.0
	Hispanic	197	30.5	8.6	60.9
	American Indian	9	44.4	22.2	33.3
	Multiracial	60	21.7	16.7	61.7
	Native Hawaiian/Pacific Islander	1			100.0
	White	437	31.4	10.1	58.6
	All	1097	31.5	10.3	58.2
SWD	Autism	380	25.0	5.8	69.2
	Serious Emotional Disability	10	50.0	50.0	
	Hearing Impairment	3	33.3		66.7
	Intellectual Disability - Mild	182	50.6	24.2	25.3
	Intellectual Disability - Moderate	289	36.0	5.5	58.5
	Intellectual Disability - Severe	41	4.9	2.4	92.7
	Specific Learning Disability	9	55.6	33.3	11.1
	Multiple Disability	96	15.6	2.1	82.3
	Other Health Impairment	50	32.0	28.0	40.0
	Orthopedic Impairment	2			100.0
	Other	16	31.3	12.5	56.3
	Traumatic Brain Injury	17	23.5	23.5	52.9
	Visual Impairment	2	100.0		
	All	1,097	31.5	10.3	58.2
EDS	Not Economically Disadvantaged	564	27.8	8.0	64.2
	Economically Disadvantaged	533	35.5	12.8	51.8
	All	1097	31.5	10.3	58.2
ELs	Regular	955	31.6	10.8	57.6
	Other	3			100.0
	English Language Learner	139	31.7	7.2	61.2
	All	1097	31.5	10.3	58.2

Note: Not Proficient=Inconsistent Understanding, Not CCR 3= Sufficient Understanding, Not CCR 4= Thorough Understanding, on Track for Competitive Employment, CCR

Table 6. 2020-21 NCEXTEND1 Grade 8 Reading Proficiency Classifications by Subgroups

Category	Sub-Category	N	Not Proficient	Level 3	Level 4
Ethnicity	Asian	36	13.9	2.8	83.3
	Black	400	36.5	7.3	56.3
	Hispanic	193	26.4	2.6	71.0
	American Indian	10	20.0	40.0	40.0
	Multiracial	58	29.3	12.1	58.6
	Native Hawaiian/Pacific Islander	3			100.0
	White	526	34.8	7.2	58.0
	All	1226	33.0	6.9	60.2
SWD	Autism	433	29.8	4.6	65.6
	Deaf-Blindness	2	50.0		50.0
	Deafness	2			100.0
	Serious Emotional Disability	2	50.0		50.0
	Hearing Impairment	1			100.0
	Intellectual Disability - Mild	184	54.4	15.8	29.9
	Intellectual Disability - Moderate	306	37.6	3.9	58.5
	Intellectual Disability - Severe	51	5.9		94.1
	Specific Learning Disability	1	100.0		
	Multiple Disability	139	14.4	4.3	81.3
	Other Health Impairment	62	33.9	21.0	45.2
	Orthopedic Impairment	1	100.0		
	Other	21	38.1		61.9
	Traumatic Brain Injury	18	16.7	16.7	66.7
	Visual Impairment	3	33.3	33.3	33.3
	All	1226	33.0	6.9	60.2
EDS	Not Economically Disadvantaged	650	31.4	6.5	62.2
	Economically Disadvantaged	576	34.7	7.3	58.0
	All	1226	33.0	6.9	60.2
ELs	Regular	1101	33.5	7.3	59.2
	Other	1			100.0
	English Language Learner	124	28.2	3.2	68.6
	All	1226	33.0	6.9	60.2

Note: Not Proficient=Inconsistent Understanding, Not CCR 3= Sufficient Understanding, Not CCR 4= Thorough Understanding, on Track for Competitive Employment, CCR

Table 7. 2020-21 NCEXTEND1 English II at Grade 10 Proficiency Classifications by Subgroups

Category	Sub-Category	N	Not Proficient	Level 3	Level 4
Ethnicity	Asian	15	40.0	6.7	53.3
	Black	273	33.7	3.7	62.6
	Hispanic	138	23.9	5.1	71.0
	American Indian	14	28.6	14.3	57.1
	Multiracial	40	35.0	2.5	62.5
	Native Hawaiian/Pacific Islander	1	100.0		
	White	391	33.5	5.9	60.6
	All	872	32.2	5.1	62.7
SWD	Autism	288	29.9	4.2	66.0
	Deaf-Blindness	2	50.0		50.0
	Serious Emotional Disability	2	50.0		50.0
	Intellectual Disability - Mild	127	54.3	11.0	34.7
	Intellectual Disability - Moderate	246	26.8	4.1	69.1
	Intellectual Disability - Severe	45	11.1	2.2	86.7
	Specific Learning Disability	5	80.0	20.0	
	Multiple Disability	90	23.3		76.7
	Other Health Impairment	44	29.6	13.6	56.8
	Orthopedic Impairment	2	50.0		50.0
	Other	10	60.0		40.0
	Traumatic Brain Injury	10	70.0		30.0
	Visual Impairment	1	100.0		
	All	872	32.2	5.1	62.7
EDS	Not Economically Disadvantaged	520	31.9	4.8	63.3
	Economically Disadvantaged	352	32.7	5.4	61.9
	All	872	32.2	5.1	62.7
ELs	Regular	785	32.9	5.0	62.2
	Other	4			100.0
	English Language Learner	83	27.7	6.0	66.3
	All	872	32.2	5.1	62.7

Note: Not Proficient=Inconsistent Understanding, Not CCR 3= Sufficient Understanding, Not CCR 4= Thorough Understanding, on Track for Competitive Employment, CCR

Table 8. 2020-21 NCEXTEND1 Grade 5 Science Proficiency Classifications by Subgroups

Category	Sub-Category	N	Not Proficient	Level 3	Level 4
Ethnicity	Asian	44	27.3	18.2	54.6
	Black	378	36.2	25.4	38.4
	Hispanic	207	41.1	12.6	46.4
	American Indian	15	40.0	33.3	26.7
	Multiracial	53	37.7	24.5	37.7
	Native Hawaiian/Pacific Islander	4	25.0		75.0
	White	433	35.3	19.2	45.5
	All	1134	36.5	20.4	43.1
SWD	Autism	457	34.8	21.4	43.8
	Deaf-Blindness	2			100.0
	Serious Emotional Disability	4		100.0	
	Hearing Impairment	1	100.0		
	Intellectual Disability - Mild	167	51.5	37.7	10.8
	Intellectual Disability - Moderate	253	37.2	9.9	53.0
	Intellectual Disability - Severe	40	20.0	10.0	70.0
	Specific Learning Disability	8	37.5	50.0	12.5
	Multiple Disability	128	29.7	7.8	62.5
	Other Health Impairment	42	35.7	38.1	26.2
	Orthopedic Impairment	7	71.4		28.6
	Other	15	26.7	26.7	46.7
	Traumatic Brain Injury	9		33.3	66.7
	Visual Impairment	1	100.0		
	All	1134	36.5	20.4	43.1
EDS	Not Economically Disadvantaged	583	35.5	17.5	47.0
	Economically Disadvantaged	551	37.6	23.4	39.0
	All	1134	36.5	20.4	43.1
ELs	Regular	974	36.1	21.9	42.0
	Other	3			100.0
	English Language Learner	157	39.5	11.5	49.0
	All	1134	36.5	20.4	43.1

Note: Not Proficient=Inconsistent Understanding, Not CCR 3= Sufficient Understanding, Not CCR 4= Thorough Understanding, on Track for Competitive Employment, CCR

Table 9. 2020-21 NCEXTEND1 Grade 8 Science Proficiency Classifications by Subgroups

Category	Sub-Category	N	Not Proficient	Level 3	Level 4
Ethnicity	Asian	36	36.1	8.3	55.6
	Black	394	43.7	19.0	37.3
	Hispanic	190	40.0	11.1	49.0
	American Indian	10	40.0	40.0	20.0
	Multiracial	58	44.8	17.2	37.9
	Native Hawaiian/Pacific Islander	3			100.0
	White	521	45.1	17.7	37.2
	All	1212	43.4	16.9	39.7
SWD	Autism	426	38.5	15.7	45.8
	Deaf-Blindness	1	100.0		
	Deafness	2			100.0
	Serious Emotional Disability	2	50.0		50.0
	Hearing Impairment	1			100.0
	Intellectual Disability - Mild	182	55.0	34.6	10.4
	Intellectual Disability - Moderate	304	51.6	12.8	35.5
	Intellectual Disability - Severe	51	25.5		74.5
	Specific Learning Disability	1	100.0		
	Multiple Disability	138	31.2	5.8	63.0
	Other Health Impairment	62	43.6	32.3	24.2
	Orthopedic Impairment	1	100.0		
	Other	20	50.0	10.0	40.0
	Traumatic Brain Injury	18	38.9	22.2	38.9
	Visual Impairment	3	33.3	66.7	
	All	1212	43.4	16.9	39.7
EDS	Not Economically Disadvantaged	640	43.0	15.6	41.4
	Economically Disadvantaged	572	43.9	18.4	37.8
	All	1212	43.4	16.9	39.7
ELs	Regular	1089	43.4	17.5	39.0
	Other	1			100.0
	English Language Learner	122	43.4	11.5	45.1
	All	1212	43.4	16.9	39.7

Note: Not Proficient=Inconsistent Understanding, Not CCR 3= Sufficient Understanding, Not CCR 4= Thorough Understanding, on Track for Competitive Employment, CCR

Table 10. 2020-21 NCEXTEND1 Biology at Grade 10 Proficiency Classifications by Subgroups

Category	Sub-Category	N	Not Proficient	Level 3	Level 4
Ethnicity	Asian	15	33.3	13.3	53.3
	Black	275	45.1	11.3	43.6
	Hispanic	138	46.4	9.4	44.2
	American Indian	14	35.7	28.6	35.7
	Multiracial	41	39.0	14.6	46.3
	Native Hawaiian/Pacific Islander	1		100.0	
	White	390	47.2	13.1	39.7
	All	874	45.5	12.4	42.1
SWD	Autism	288	45.1	11.1	43.8
	Deaf-Blindness	2		50.0	50.0
	Serious Emotional Disability	2	100.0		
	Intellectual Disability - Mild	127	66.1	21.3	12.6
	Intellectual Disability - Moderate	247	45.8	9.7	44.5
	Intellectual Disability - Severe	46	17.4	6.5	76.1
	Specific Learning Disability	4		100.0	
	Multiple Disability	92	31.5	3.3	65.2
	Other Health Impairment	44	43.2	22.7	34.1
	Orthopedic Impairment	2		50.0	50.0
	Other	9	55.6	22.2	22.2
	Traumatic Brain Injury	10	80.0		20.0
	Visual Impairment	1		100.0	
	All	874	45.5	12.4	42.1
EDS	Not Economically Disadvantaged	520	44.0	11.9	44.0
	Economically Disadvantaged	354	47.7	13.0	39.3
	All	874	45.5	12.4	42.1
ELs	Regular	785	45.4	12.4	42.3
	Other	4	50.0		50.0
	English Language Learner	85	47.1	12.9	40.0
	All	874	45.5	12.4	42.1

Note: Not Proficient=Inconsistent Understanding, Not CCR 3= Sufficient Understanding, Not CCR 4= Thorough Understanding, on Track for Competitive Employment, CCR

Appendix 8–E

WinScan Score Interpretive guide

Interpretive Guide to the North Carolina Winscan Score Reports (nc.gov)